

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 22, 2004, 21:22:24 ; Search time 5080 Seconds
(without alignments)
4873.179 Million cell updates/sec

Title: US-09-916-849A-1
Perfect score: 4369
Sequence: 1 MGLPRLASLLLLQVWLQ.....NEWGSRFKKLADYGGGEDD 829

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 550265578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame_plus_p2n.model -DRV=xlh
-Q=/cg2_1/USPTO.spool_h/US09916849/runat_21092004_145803_7018/app_query.fasta_1.967
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09916849 -CGN_1_1_3609@runat_21092004_145803_7018 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hctc:*
9: gb_estl:*
10: gb_est2:*
11: gb_hctc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_esthum:*
18: em_estinv:*
19: em_estpln:*
20: em_estvrt:*
21: em_estfun:*
22: em_estmam:*
23: em_estmus:*
24: em_estpro:*
25: em_estrod:*
26: em_estphg:*
27: em_estvrl:*
28: gb_gestl:*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4310	98.6	3632	11	BC014462	BC014462 Homo sapi
2	3564.5	81.6	3971	11	AK031265	AK031265 Mus muscu
3	3560	81.5	3233	11	AK045041	AK045041 Mus muscu
4	2351	53.8	4415	11	AK076369	AK076369 Mus muscu
5	1701	38.9	3136	11	AK032826	AK032826 Mus muscu
6	1486	34.0	1016	12	BM923882	BM923882 AGENCOURT
7	1459	33.4	873	14	CA453911	CA453911 AGENCOURT
8	1440	33.0	887	14	CA455046	CA455046 AGENCOURT
9	1417.5	32.4	1049	13	BQ936686	BQ936686 AGENCOURT
10	1408	32.2	912	13	BQ927154	BQ927154 AGENCOURT
11	1382.5	31.6	846	13	BQ690885	BQ690885 AGENCOURT
12	1382.5	31.6	918	14	CA454560	CA454560 AGENCOURT
13	1379.5	31.6	912	13	BQ928283	BQ928283 AGENCOURT
14	1350	30.9	875	14	CA489015	CA489015 AGENCOURT
15	1343	30.7	913	13	BQ676796	BQ676796 AGENCOURT
16	1342	30.7	836	12	BM045728	BM045728 603624563
17	1329	30.4	865	13	BU543746	BU543746 AGENCOURT
18	1324.5	30.3	999	12	BM804940	BM804940 AGENCOURT
19	1321	30.2	892	14	CA489338	CA489338 AGENCOURT
20	1312	30.0	894	14	CA489124	CA489124 AGENCOURT
21	1310.5	30.0	899	14	CA454360	CA454360 AGENCOURT
22	1302	29.8	870	10	BE378969	BE378969 601237596
23	1293.5	29.6	773	10	BE740340	BE740340 601594641
24	1277	29.2	1122	12	BM545620	BM545620 AGENCOURT
25	1268	29.0	958	13	BQ886447	BQ886447 AGENCOURT
26	1264.5	28.9	927	14	CA489357	CA489357 AGENCOURT
27	1261	28.9	1001	13	BQ072354	BQ072354 AGENCOURT
28	1250.5	28.6	888	14	CA487856	CA487856 AGENCOURT
29	1238	28.3	745	14	CB240552	CB240552 UI-CF-FN0
30	1231	28.2	727	12	BM047867	BM047867 603618920
31	1219.5	27.9	769	12	BG403092	BG403092 602418911
32	1219.5	27.9	778	12	BM049731	BM049731 603624246
33	1212	27.7	901	14	CA454213	CA454213 AGENCOURT
34	1212	27.7	1006	13	BU168948	BU168948 AGENCOURT
35	1210.5	27.7	916	13	BU179183	BU179183 AGENCOURT
36	1206	27.6	703	12	BM715802	BM715802 UI-E-BJ0-
37	1199	27.4	1040	12	BM909817	BM909817 AGENCOURT
38	1192	27.3	880	14	CB992509	CB992509 AGENCOURT
39	1172.5	26.8	835	14	CB996355	CB996355 AGENCOURT
40	1151	26.3	761	10	BE741190	BE741190 601593814
41	1140.5	26.1	781	14	CK032635	CK032635 AGENCOURT
42	1108	25.4	949	13	BQ679513	BQ679513 AGENCOURT
43	1104.5	25.3	757	14	CK128630	CK128630 AGENCOURT
44	1103.5	25.3	734	14	CD372275	CD372275 UI-R-G00-
45	1086.5	24.9	906	13	BU145465	BU145465 AGENCOURT

ALIGNMENTS

RESULT 1
BC014462
LOCUS BC014462 Homo sapiens cadherin 3, type 1, P-cadherin (placental), mRNA (cdna) 3632 bp mRNA linear HTC 19-NOV-2003
DEFINITION Clone IMAGE:4870356), containing frame-shift errors.
ACCESSION BC014462
VERSION BC014462.1 GI:15680220
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3632)

ORIGIN

Db 1408 GCAGTGGGCCATGAGGTGACAGGCTGACGGTCACTGATCTGACAGCGCCCAACTCACCA 1467
 Qy 361 AlaTrpArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
 Db 1468 GCGTGGCGTCCACCTACCTTATCATGGCGGTGACGACGGGACCATTTTACCATCAC 1527
 Qy 381 ThrHisProGlnSerAsnGlnGlyIleLeuThrThrArgGlyGlyLeuAspPheGluAla 400
 Db 1528 ACCACCTGAGAGCAACCAAGGCACTCTGACAAACCAAGGAGGTTTGGATTGTGAGGCC 1587
 Qy 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
 Db 1588 AAAAACCAAGCACACCTGTAGTTGAGTGACCAACAGAGGCCCTTTTGTGCTGAAGCTC 1647
 Qy 421 ProThrSerThrAlaThrIleValHisValGluAspValAsnGluAlaProValPhe 440
 Db 1648 CCAACCTCCACAGCCACCTAGTGGTCCAGCTGGAGGATGTGAATGAGGACCTGTGTGTT 1707
 Qy 441 ValProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
 Db 1708 GTCCCACTCCAAAGTCGTTGAGGTCAGAGGGCATCCCACTGGGAGGCTGTGTGT 1767
 Qy 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
 Db 1768 GTCTACACTGCAGAGACCCCTGCAAGAGGAAATCAAAAGATCAGCTACCGCATCTCTGACA 1827
 Qy 481 AspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
 Db 1828 GACCCAGAGGTGGTTCAGCATGGACCCAGAGCTGGGAGGTACAGCTGTGGGCACC 1887
 Qy 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
 Db 1888 CTCGACCGTGAGGATGAGCATGTTGTGAGGAACAACATCTATGAAGTCATGTCTGGCC 1947
 Qy 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuLeuLeuLeu 540
 Db 1948 ATGGAACAATGGAGGCCCTCCACCATGCGACGGGAACCTTCTGCTAACACTGATGAT 2007
 Qy 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
 Db 2008 GTCAACAGCATGGCCCGCTGCTGAGCCCGCTGAGTCAACATCAGTCAACCAAGCCCT 2067
 Qy 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
 Db 2068 GTGCGCCAGGTGTGAACATCAAGCAAGGACCTGTCTCCCAACACCTCCCTTTCCAG 2127
 Qy 581 AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGlyAsp 600
 Db 2128 GCCAGCTCACAGATGACTCAGACATCTACTGACGGCAGAGGTCAACGAGGAGGTGAC 2187
 Qy 601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
 Db 2188 ACAGTGCTCTGTCCCTGAAGAAGTTCCTGGAAGCAGGATACATATGACGTGCACCTTCT 2247
 Qy 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
 Db 2248 CTGTCTGACCATGGCAACAAAGACGACGTGACGGGTGATCAGGGCCACGTGTGCGACATGC 2307
 Qy 641 HisGlyHisValGluThrCysProGlyProTyrPheLeuGlyGlyPheIleLeuProValLeu 660
 Db 2308 CATGGCCATGTGAAACCTGCCTGACCTGGACCTGGAAAGAGGTTTCTATCTCTCCTGTGTG 2367
 Qy 661 GlyAlaValLeuAlaLeuLeuPheLeuLeuLeuValLeuLeuLeuValArgLysLys 680
 Db 2368 GGGGTGTCTCTGTCTGTCT 2427
 Qy 681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
 Db 2428 CGAGAGATCAAGAGGCCCT 2487
 Qy 701 TyrGlyGluGlyGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
 Db 2488 TATGGCGAAGAGGGGGTGGCGAAGAGGACCAAGGACTATGACATCACCCAGCTCCACCGA 2547

Qy 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
 Db 2548 GGTCTGGAGGCCAGGCCGAGGTGTTCTCCGCAATGACGTGGCACCAACCATCATCCCG 2607
 Qy 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
 Db 2608 ACACCCATGTACCGTCTTAGGCCACCAACCCAGATGAATCGGCAACTTTATATTGA- 2666
 Qy 760 ----- 760
 Db 2667 GGGAGGGGAGAGAGGGGCTCACAGAGAGAAATGGAGGCTTGACGTGGCAAGAGGAG 2726
 Qy 761 -----AsnLeuLysAlaAlaAsnThr 767
 Db 2727 GACCGCGCGCTCCTAACTACCTGTTCTGTGTGCGCAGAAACCTCAAGCGCGCTAAACACA 2786
 Qy 768 AspProThrAlaProProTyrAspThrLeuLeuValPheAspTyrGlyGlySerGlySer 787
 Db 2787 GACCCACAGCCCGCCCTTACGACACCTCTTGTGTGTGACTATGAGGGCAGCGGTCC 2846
 Qy 788 AspAlaAlaSerLeuSerSerLeuThrSerSerAlaSerAspGlnAspGlnAspTyrAsp 807
 Db 2847 GACGCGCGTCCCTGAGTCCCTCACCTCTCCGCTCCGACCAAGACCAAGATTACGAT 2906
 Qy 808 TyrLeuAsnGluTrpGlySerArgPheLysLysLeuAlaAspMetTyrGlyGlyGlu 827
 Db 2907 TATCTGAACGAGTGGGCGAGCCGCTTCAAGAAGTGGCAGACATGTACGTTGGCGGGAG 2966
 Qy 828 AspAsp 829
 Db 2967 GACGAC 2972
 RESULT 2
 AK031265
 LOCUS
 DEFINITION Mus musculus 13 days embryo forelimb cDNA, RIKEN full-length enriched library, clone:5930439E13 product:cadherin 3, full insert sequence.
 ACCESSION AK031265
 VERSION AK031265.1 GI:26327166
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999)
 JOURNAL 99279253
 MEDLINE 10349636
 PUBMED
 REFERENCE 2 Carninci, P. and Hayashizaki, Y. Preparation of full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)
 JOURNAL 20499374
 MEDLINE 11042159
 PUBMED
 REFERENCE 3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000)
 JOURNAL 20530913
 MEDLINE 11076861
 PUBMED

4 THE RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

6 (bases 1 to 3971)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

Direct Submission
Submitted (16-JUN-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp).
URL: <http://genome.gsc.riken.go.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: <http://genome.gsc.riken.go.jp/>
URL: <http://phantom.gsc.riken.go.jp/>

Location/Qualifiers
1. 3971
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM.DB:5930439E13"
/db_xref="MGI:2394470"
/db_xref="taxon:10090"
/clone="5930439E13"
/tissue type="forelimb"
/clone lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="13 days embryo"
48. 2516
/note="unnamed protein product: cadherin 3 (MGD)|MGI:88356, GB|X06340, evidence: BLASTN, 99%, match=3186)
putative"
/codon_start=1
/protein_id="BAC27327.1"
/db_xref="GI:26327167"
/translation="WELLSGPHAFLLLLQWLRSVSEPEYRAGFIEAGVTLVEVG
TDLFSQVLKGVKALQCGMHADNDGIIMLRGTQVGGKDAHSPPELIRRRKEWV
MPPIFVPEKNGKFFQRLNQLKSNDRGTGKI FYSTIGPGADSPEGVFTIEKSGWLL
LMPDLREKIVELYELHVSNGSEVPEPNISIIIVTDQNDKPKFTQDITFRSGVLE
GMPGTSSVMQVATDEDAVNTYNGVAYSISQEPKPHDLMTIHKSTQISVIS
GHDREKLVPRYLTQATMDGSGSTTAAVQVLNDANDNAPEFEPQKYEAWEPEVEV
GHEKORLVTDLDVNSPAMRATYHVGDDGDHFTITHPETNQGVLTITKGLDFA
QDQHTLYEVNTNEAPFAVKLPATATVYVHVKNVAPVFPVPSKVI EAQEGISIGEL
VCLYTAQDREKQDKISVTISPDANMLAVPDQSQTIAAGILDRSDRQFVKNVYEV
MYLTADSGNPPTIGTLLLLTLDINDHGPPEPQIILINQSPVPOVLNITDKLSP
NSRPFQAQTHSDDIYVMAVSEKGDFTVALSKFLKPDQTLVDLHLSLSDHNRQLTM
INATPQCDCHGVFNDCFPWKGDFILPILGAVLALLTLLALLLLHKKRKKPKPLLL
PEDDTDRNVYVYGEESGDEQDQITQLHRLGEARPEVLRNDVVPFTIPTRPRR
PANPDEIGNFIIENKAANDTPTAPYDLSLIVFDYEGSGSDAASLSLTTSASQDQD

CDS

```

QY 299 GluAlaThrAspMetAspGlyAspGlySerThrThrAlaValAlaValGluIle 318
Db 918 CAGCCACAGACATGGATGAGAGGGCTCTACCCAGCAGCAGAGCCGTTGTCAAATC 977
QY 319 LeuAspAlaAsnAspAsnAlaProMetPheAspProGlnIleTyrGluAlaHisValPro 338
Db 978 CTTGATGCCAACGATACCGCTCCGGAGTTTGAGCCGAGAGATATCAGAGCTTGGGTGCCT 1037
QY 339 GluAsnAlaValGlyHisGluValGlnAlaArgLeuThrValThrAspLeuAspAlaProAsn 358
Db 1038 GAGAACGAGATGGGCCATGAGGTACAGAGGCTCAGCAGTCACTGATCTCGATGTCCCAAC 1097
QY 359 SerProAlaTirpArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThr 378
Db 1098 TCGCCAGCGTGGGTGCCACTACCATCTGCTGGAGGTGATGATGGGAGCACTTCAC 1157
QY 379 IleThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPhe 398
Db 1158 ATCACCACCTCACCAGAGACCAACCAAGGCTCCTGACACCAACCAAGAGGTTTGGATTTT 1217
QY 399 GluAlaLysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeu 418
Db 1218 GAGGCTCAGGACCAACACACCCCTGTATGTAGAGTACCACAGGCTCCCTTTGCAGTG 1277
QY 419 LysLeuProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaPro 438
Db 1278 AAGCTCCGACAGCCACTGCCACCGTGGTGTCTCATGTGAAGATGTCAACGAAGCCCT 1337
QY 439 ValPheValProSerLysValValGluValGlnGluGlyIleProThrGlyGluPro 458
Db 1338 GTGTTTGTCCACCTTCCAAAGGTCATGTAGGCGCCAGGAAGGCATCTCTATTGGGGAATG 1397
QY 459 ValCysValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIle 478
Db 1398 GTCTGCATCTATACCGACAGGACCCAGACAGGAGGACGAGAGATCAGCTACACCATC 1457
QY 479 LeuArgAspProAlaGlyTirpLeuAlaMetAspProAspSerGlyGlnValThrAlaVal 498
Db 1458 TCAGAGATGCCAGCAACTGGCTGTGTGGAGCCAGACAGTGGTCAGATACTGCCGCA 1517
QY 499 GlyThrLeuAspArgGluAspGluGlnPheValArgAsnIleTyrGluValMetVal 518
Db 1518 GGCATCTTGATCTGAGACGACGAGCTTTGTGAAACAACATGTACGAAGTCATGGTT 1577
QY 519 LeuAlaMetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuThrLeu 538
Db 1578 TTGGCCACAGACAGTGGAAACCCCTCCACACCGGCACCTGGGACCTCTGCTTACACTT 1637
QY 539 IleAspValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGln 558
Db 1638 ACTGATCAACAGACCATGGCCGATCCCTGTGAACCCAGGACATCATCTGTAAACCAA 1697
QY 559 SerProValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerPro 578
Db 1698 AGCCCTGTGCCTCAAGTGTGAACATCACTGACAGGACCTGTCCCCCACTCTCCCT 1757
QY 579 PheGlnAlaGlnLeuThrAspAspSerAspIleTyrThrAlaGluValAsnGluGlu 598
Db 1758 TTCAGGCCCGAGTAAACATGACTCAGATATCTACTGGATGGCAGAGTACGAGCAAA 1817
QY 599 GlyAspThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHis 618
Db 1818 GGNAGACCGTGGCTTGTCCCTGAGAGAGTTCCTGAAACAAGACACGATGACTGTGAT 1877
QY 619 LeuSerLeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCys 638
Db 1878 CTTCTCTTGTCTGACCATGCAACAGGAAAGAGTAAACCATGATCAGGCCACTGTGTGT 1937
QY 639 AspCysHisGlyHisVal---GluThrCysProGlyProTirpLysGlyGlyPheLeuLeu 657
Db 1938 GACTGCCATGGCCAAAGTGTTCATGACTGCCCAACCCCTGGAAGGCTGTTCATCCTC 1997
QY 658 ProValLeuGlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuVal 677

```

```

Db 1998 CCCATCTGGGTGCTCTGCTGGCACTGCTGACCCCTTACTGGCACTCCTCTGTGTGTG 2057
QY 678 ArgLysLysArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsn 697
Db 2058 AGGAAGAAGAGAGAGGTCAAAGAGCCCTTCTGCTCCAGAGATGACACCGGAGACAA 2117
QY 698 ValPheTyrTyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGln 717
Db 2118 GTCTTCTATTATGGAGAAGAGGGTGGTGGTGAAGAGGACCAAGGACTATGACATCACC 2177
QY 718 LeuHisArgGlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThr 737
Db 2178 CTCACCGGGGACTGGAGCCAGGCTGAGGTGTTCTCCGAACGATGTAGTCCCAACC 2237
QY 738 IleIleProThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPhe 757
Db 2238 TTCATCCCAACCCCATGTACCGACCCCGGCCCAACCCAGATGAATCGGAATCTTC 2297
QY 758 IleIleGluAsnLeuLysAlaAlaAsnThrAspProThrAlaProTyrAspThrLeu 777
Db 2298 ATCATCGAAGACCTGAAGCTGCCAACAACGACCTACTGCCCGCCCTACGACTCCCTG 2357
QY 778 LeuValPheAspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSer 797
Db 2358 CTGGTTTTTGACTACGAGGGCAGCGGCTCTGATCGGCTCCCTGAGCTCCCTCACCACC 2417
QY 798 SerAlaSerAspGlnAspGlnAspTyrAspTyrIleAsnGluThrGlySerArgPheLys 817
Db 2418 TCCGCTCCGACCCAGGATCAGGACTCAACAACCTTAACGAGTGGGAAGTCGATTCAAG 2477
QY 818 LysLeuAlaAspMetTyrGlyGlyGlyLeuAspAsp 829
Db 2478 AAACCTGGCGGACATGTATGTTGGCGGTGAGGATGAC 2513

RESULT 3
AK045041 3233 bp mRNA linear HTC 20-SEP-2003
LOCUS Mus musculus 9.5 days embryo parthenogenote cDNA, RIKEN full-length
DEFINITION enriched library, clone:BI30021C20 product:cadherin 3, full insert
sequence.
ACCESSION AK045041
VERSION AK045041.1 GI:26337020
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
2 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Tawara, Y., Izawa, M., Ohara, E., Watabiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuda, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)

```


Db 818 AGCCAAGAGCGAAGAGGCCACACGACCTCATGTTCACCATCCATATAAAGACGCGGAACC 877
 Qy IleserValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIle 298
 Db 878 ATTAGCGTCATATCCAGTGGCGTGGACCGAGAGAAAGTCCCTGAGTACACACTGACCGTC 937
 Qy GlnIleThrAspMetAspGlyAspGlySerThrThrAlaValAlaValGluIle 318
 Db 938 CAGGCCACAGACATGGAATGGAGAGGGCTCTACCAACGACGCGCAGAGCGCGTGTGTGCAATC 997
 Qy LeuAspAlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValPro 338
 Db 998 CTGTGATCCCAACGATACGCTCCGAGTGTGAGCCGACGAGATATGAGGCTTGGTGCCCT 1057
 Qy GluAsnAlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsn 358
 Db 1058 GAGAACCAAGTGGGCCATGAGGTACAGAGGCTGACAGGCTGACAGTCTGATCTCGATGTCCCAAC 1117
 Qy SerProAlaIleArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThr 378
 Db 1118 TCGCCAGCGTGGGTGCCACTACCATCTGTGGAGGATGATGAGGGACCATTTCCACC 1177
 Qy IleThrThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPhe 398
 Db 1178 ATCACCACCTCACCAGACCAACCAAGGCGTCTGACAAACCAAGAGGCTTTTGGATTT 1237
 Qy GluAlaLysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeu 418
 Db 1238 GAGCTCAGGACCAACACACCCCTGTATGTAGAAGTACCAACAGAGCTCCCTTTGCGATG 1297
 Qy LysLeuProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaPro 438
 Db 1298 AAGCTCCGACAGCCACTGCCCGTGGTGTCCATGTGAAGATGTCAACGAGGCCCT 1357
 Qy ValPheValProSerLysValValGluValGlnGluGlyIleProThrGlyGluPro 458
 Db 1358 GTGTTTGTTCACCTTCCAGAGTCAATGAGCGCCAGGAAGGCAATCTCTATTGGGAATG 1417
 Qy ValCysValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIle 478
 Db 1418 GTCTGATCTATACCGCACAGACCCAGACAGAGGAGCAGGAAGATCAGCTACACCATC 1477
 Qy LeuArgAspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaVal 498
 Db 1478 TCGAGAGATCCAGCCAACTGGCTGTGTGACCCAGACAGTGGTCAGATACTGCCGCA 1537
 Qy GlyThrLeuAspArgGluAspGluGlnPheValArgAsnIleTyrGluValMetVal 518
 Db 1538 GGCATCTTGGATCGTGAGGACGAGCAGTGTGTGAAAAACAATGTCTACGAAGTCAATGGTT 1597
 Qy LeuAlaMetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeu 538
 Db 1598 TTGGCCACAGACAGTGGAAACCTCCACACCGGCACTGGGACCTCTCTGTATACATT 1657
 Qy IleAspValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGln 558
 Db 1658 ACTGATCAACAGCAGATGCCCGATCCCTGGAACCCAGGACAGATCATCATCTGTAAACAA 1717
 Qy SerProValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerPro 578
 Db 1718 AGCCCTGTGCTCAAGTGTGAACATCATGACAGGACCTGTCCCCCACTCTCCCT 1777
 Qy PheGlnAlaGlnLeuThrAspSerAspIleTyrTrpThrAlaGluValAsnGluGlu 598
 Db 1778 TTCAGGCCCGCAGTAAACATGACTCAGATATCTACTGGATGGCAGAGTACAGAGAA 1837
 Qy GlyAspThrValValLeuSerLeuLysPheLeuLysGlnAspThrTyrAspValHis 618
 Db 1838 GGAGACACCGTGGCTTGTCCCTGGAAGAAGTTCCTGAAACAAGACAGCATGATGATGAT 1897
 Qy LeuSerLeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCys 638
 Db 1898 CTTTCTTTCTGTGACCATGCAACAGGGAACAGCTAACCATCATCAGGGCCACTGTGTGT 1957

Qy 639 AspCysHisGlyHisVal---GluThrCysProGlyProTyrIleGlyGlyPheIleLeu 657
 Db 1958 GACTGCCATGGCAAGTGTCAATGACTGCCCGACAGCCCTGGAAGGTTGGTTTCATCCTC 2017
 Qy 658 ProValLeuGlyAlaValLeuAlaLeuPheLeuLeuValLeuLeuLeuVal 677
 Db 2018 CCCATCTGGTGTCTCTGGCACTGTGTACCCCTTCTACTGCGCACTCTCTCTGTGGTG 2077
 Qy 678 ArgLysArgGlyIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsn 697
 Db 2078 AGAAGAGAGAGAAAGTCAAGAGCCCTTCTGCTCCAGAGAGATGACACGCGAGACAT 2137
 Qy 698 ValPheTyrTyrGlyGluGluGlyGlyGluGluAspGlnAspTyrAspIleThrGln 717
 Db 2138 GTCTTCTATTATGAGAGAGAGGCTGTGGTGAAGAGGACCAAGACTATGACATCACCAA 2197
 Qy 718 LeuHisArgGlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThr 737
 Db 2198 CTCACCGGGGACTCGAGCGCCAGGCTGAGTGGTCTCCGAAACGATGTATGCAACC 2257
 Qy 738 IleIleProThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPhe 757
 Db 2258 TTCATCCCAACCCCATGTACCGACCCCGCGCCCAACCCAGATGAATCGGGAATTC 2317
 Qy 758 IleIleGluAsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeu 777
 Db 2318 ATCATCGAAGAACTGAAGGCTGCCAACACTGACCTACTGCGCCGCCCTAGCACTCCCTG 2377
 Qy 778 LeuValPheAspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSer 797
 Db 2378 CTGGTTTTTGACTACGAGGGCAGCGGCTCTGATCGCGCTCCCTGAGCTCCCTCACCAC 2437
 Qy 798 SerAlaSerAspGlnAspGlnAspTyrAspTyrLeuAsnGluTyrGlySerArgPheLys 817
 Db 2438 TCGCGCTCCGACCAAGATCAGGACTACAACTACCTTACGAGTGGGGAAGTCGATTCAG 2497
 Qy 818 LysLeuAlaAspMetTyrGlyGlyGlyGluAspAsp 829
 Db 2498 AAATGGCGGACATGTATGTGGCGGTGAGGATGAC 2533
 RESULT 4
 Locus AK076369 4415 bp mRNA linear HTC 18-SEP-2003
 DEFINITION Mus musculus 10 days neonate skin cDNA, RIKEN full-length enriched library, clone:473469F18 product:cadherin 1, full insert sequence.
 ACCESSION AK076369
 VERSION AK076369.1 GI:26096766
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1
 AUTHORS Carninci,P. and Hayashizaki,Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636
 REFERENCE 2
 AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20499374
 PUBMED 11042159
 REFERENCE 3
 AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kiteunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,

Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multipillar sequencer
Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL
MEDLINE
20530913
PUBMED
11076861

REFERENCE

4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.

Functional annotation of a full-length mouse cDNA collection

Nature 409, 685-690 (2001)

REFERENCE

5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

Analysis of the mouse transcriptome based on functional annotation

of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

REFERENCE

6 (bases 1 to 4415)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Horii, F., Inotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Oheato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toyota, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

Direct Submission

Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: http://genome.gsc.riken.go.jp/

Location/Qualifiers

FEATURES

source

1..4415

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="PANTOM:DB:4732469F18"

/db_xref="MGI:2350926"

/clone="caxon:10090"

/tissue type="skin"

/clone_lib="RIKEN full-length enriched mouse cDNA library"

/dev stages="10 days neonate"

129..2784

/note="Caderin 1 (MGI:88354, GB|NM_009864, evidence: BLASTN, 99%, match=4414)

putative"

4393..4398

/note="putative"

4415

/note="putative"

ORIGIN

Alignment Scores:

Pred. No.: 8,33e-238 Length: 4415

Score: 2351.00 Matches: 469

Percent Similarity: 67.01% Conservative: 122

Best Local Similarity: 53.17% Mismatches: 224
Query Match: 53.81% Indels: 67
DB: 11 Gaps: 13
US-09-916-849a-1 (1-829) x AK076369 (1-4415)
QY 9 AlaSerLeuLeuLeuLeuGlnVal---CysTrpLeuGlnCysAlaLaserGluPro---- 26
156 GCGCTCTCTGCTCTGCTGCGAGTCTCTCTCATGGCT---TCCAGGAGCTGGAGCTGAG 212
QY 27 ---CysArgAlaValPheArgGluAlaGluValThrLeuGluAlaGlyAlaGluGln 45
213 TCCTGCACTCCGCGCTTCAGTTCGAGGTCTACACCTCCGCGTCCGGAGAGCACTG 272
QY 46 GluProGlyGlnAlaLeuGlyLysVal---PheMetGlyCysProGlyGlnGluProAla 64
273 GAGAGAGGCCATGCTCTGGGCGAGAGTGAAGATTTGAAGATGCACTGCGCGCCAGGACA 332
QY 65 LeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThrValGluGluArg 84
333 GCCTTCITTCGAGAGACTCCGATTCAAAGTGGCGACACGCGACCATCACAGTGAAG 392
QY 85 ArgSerLeuLys----- 88
393 CGGCATCTAAAGCTCCCAAGCTGGAGACCAAGTTTCTCTGTCGCGCGCGGAGACTCCAGT 452
QY 88 ----- 88
453 CATAGGGAGCTCTCTACCAAGTCAAGTCAAGTCCATGGGCGACCAACCATCACCGGCAC 512
QY 89 -----GluArgAsnPro---LeuLysIlePheProSerLysArg 100
513 CACCACCGCGACCTGCTCTGAATCCACCCAGAGCTGCTCATGTTCCTCCAGCTGTAC 572
QY 101 Ile---LeuArgArgHisLysArgAspTrpValValAlaProIleSerValProGluAsn 119
573 CCAGGTCTCAGACAGACAGAACCGAGACTGGGTGTCATCCCTCCCATCAGCTGCCCGGAAAT 632
QY 120 GlyLysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThr 139
633 GAAAGGGCGCAATTCCAAAGAACCTGGTTTCAGATCAATCAAAAGAGGACAGGACAAAGAA 692
QY 140 LysIlePheTrpSerIleThrGlyProGlyAlaAspSerProGluGlyValPheAla 159
693 AAGTTTTTCACGATCACCGCGCCAGAGAGTGAACAAACCCCGCTTGGCGTTTCATC 752
QY 160 ValGlnLysGluThrGlyTrpLeuLeuLeuAsnLysPheLeuAspArgGluGluLeuAla 179
753 ATTGAGAGGAGACAGGCTGGCTGAAAGTGACACAGACCTCTGGATAGAGAGCCATTGCC 812
QY 180 LysTrpGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMet 199
813 AAGTACATCTTCTTCTCTGCGGTGTCATCAATGGGAGCGGTGGAGATCCCATG 872
QY 200 AsnIleSerIleValThrAspGlnAsnAspHisLysPheProLysPheThrGlnAspThr 219
873 GAGATAGTGTATCAGTGAAGATCAGATCAAGATCAACAGGCGCAGTTTACCAGCGCGTTC 932
QY 220 PheArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAla 239
933 TTTCAGGAGTCTGTCGAGAGCGCTGTTCAGAGAACCTCCGTGATGAAGTCTCAGCC 992
QY 240 ThrAspGluAspAspAlaIleThrThrThrAsnGlyValValAlaThrSerIleHisSer 259
993 ACCGATCAGACGATGACGTCACACCTCAACAGCTGCCATCAGCTACACCATCCTCAGC 1052
QY 260 GlnGluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIle 279
1053 CAGATCTCTGAGTGTCTCAAAAACATGTTCACTGTCAATAGGACACCGGGGTCTATC 1112
QY 280 SerValIleSer-SerGlyLeuAspArgGlnLysValProGluThrThrLeuThrIleGln 299
1113 AGTGTGCTCACCTCTGGGCTGGACCGGAGAGAGTTTACCCCTACATACACTCTCTGCTGTTCA 1172

Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861

AUTHORS

The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

AUTHORS

The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, W.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.

AUTHORS

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/
Location/Qualifiers

FEATURES

source

```
1. 3136
    /organism="Mus musculus"
    /mol_type="mRNA"
    /strain="C57BL/6J"
    /db_xref="FANTOM_DB:6720458H05"
    /db_xref="MGI:2396474"
    /db_xref="taxon:10090"
    /clone="6720458H05"
    /sex="male"
    /tissue_type="olfactory duct includes surrounding region"
    /clone_lib="RIKEN full-length enriched mouse cDNA library"
    /dev_stage="12 days embryo"
    96..2816
    /note="unnamed protein product; cadherin 2 (MGD|MG1.88355,
    GB|NM_007664, evidence: BLASTN, 99%, match=3135)
    putative"
    /codon_start=1
    /protein_id="BAC28044.1"
    /db_xref="GI:26328611"

```

CDS

ORIGIN

Alignment Scores:
Pred. No.: 7,39e-169 Length: 3136
Score: 1701.00 Matches: 392
Percent Similarity: 56.41% Conservative: 127
Best Local Similarity: 42.61% Mismatches: 289
Query Match: 38.93% Indels: 112
Dbs: 11 Gaps: 29

US-09-916-849A-1 (1-829) x AK032826 (1-3136)

QY 2 GlyLeuProArgGlyProLeuAlaSerLeuLeuLeuGlnValCysTrpLeuGlnCys 21
DB 111 GGAGCGCGCGG-----ACCTGTCGCGCTTCTGGCGCGCTTCTTCAGCG 158
QY 22 AlaAlaSerGluPro-----CysArgAlaValPheArgGluAlaGluVal 36
DB 159 TCTGTGAGACTTCTGTGAAATTCATTATGCAAGTCTGATTTCTCTGAA--GATGTT 215
QY 37 ThrLeuGluAlaGlyGlyAlaGluGlnGluProGlyGlnAlaLeuGlyLysVal---Phe 55
DB 216 TAGACGCGAGTCTTACCGAGGATGTCACGAGGACAGCGCCCTTCTCAATGTGAATTC 275
QY 56 MetGlyCys-----ProGlyGlnGluProAlaLeuPheSer 67
DB 276 AGCAACTGCAATAGAAAAGAAAGTTCAGTATGAAAGCAGCGAGCAGCAGATTTCAAG 335
QY 68 ThrAspAsnAspAsp-----PheThrValArgAsn----- 77
DB 336 GTGACGAGGAGCGACCGGTGTATGCTGTGAGAAAGTTCCTCTCACTGCAGACAGGCA 395
QY 78 -----GlyGluThrValGlnGluArg----- 84
DB 396 AAGTTCCTGATATATGCCAAGACAAAGAAACCCAGGAAAGTGGCAGGTAGCTGTAAC 455
QY 85 -----ArgSerLeuLysGluArgAsnProLeuLys--- 94
DB 456 CTGACCGCGGAGCCCAACCTGACTGAGGAGCCTATGAGGAAGCAACATGAAATTTGAAGAA 515
QY 95 --IlePheProSerLys-----ArgIleLeuArgArgHisLysArgAsp 108
DB 516 ATAGTATTCCTAGACAACTTCCCAAGCAGCAGTGGAGCTCTACAAAGCAGAGAGAGAC 575
QY 109 TrpValValAlaProIleSerValProGluAsnGlyLysGlyProPheProGlnArgLeu 128
DB 576 TGGGTCTATCCCGCAATCACTTCCAGAAACCTCCAGAGAGCCCTTCTCTCAAGACCTT 635
QY 129 AsnGlnLeuLysSerAsnLysAspArgAspThrLysIlePheTyrSerIleThrGlyPro 148
DB 636 GTCGAATCAGCTTGTATAGAGATAAAACCTTTCCTCGAGATACAGCGTCACTGGGCCA 695
QY 149 GlyAlaAspSerProGluGluGlyValPheAlaValGluLysGluThrGlyTrpLeuLeu 168
DB 696 GGAGCTGACCAAGCCTCCAAACCGGCACTTCTATTATCAACCCCATCTCAGGACAGCTGCA 755
QY 169 LeuAsnLysProLeuAspArgGluIleAlaLysTyrGluLeuPheGlyHisAlaVal 188
DB 756 GTCGAAGCCTCTGGATCGAGAGCTGATACCCCGGTTTTCCTTCAGAGCAGCATGCAATG 815
QY 189 SerGluAsnGlyAlaSerValGluAspProMetAsnIleSerIleLeuThrAspGln 208
DB 816 GACATCAATGGCAATCAAGTGGAGAACCCCATTTGACATTTGATTTGACATG 875

Qy	209	AsnAspHisLysProLysPheThrGlnAspThrPheArgGlySerValLeuGluGlyVal	228
Db	876	AATGATACACACCTGAGTTTCTGCACACAGGTTTGGAAATGGGTCTGTTCCAGAGGGATCA	935
Qy	229	LeuProGlyThrSerValMetGlnValThrAlaThrAspGluAspAspAlaIleTyThr	248
Db	936	AAGCCTGGGACGTATGTGATCACGGTCACTGCCATTGATCGGATGATCCA---AATGCC	992
Qy	249	TyrAsnGlyValValAlaTySerIleHisSerGlnGluProLysAspProHisAspLeu	268
Db	993	CTGAATGGAATCGTCGGGTACAGATCTGTCCAGGCGGCCACAGACACTTACCCCAAC	1052
Qy	269	MetPheThrIleHisArgSerThrGlyThrIleSerValIleSerSerGlyLeuAspArg	288
Db	1053	ATGTTTACATCAAAATGACACTGGGACATCATCACTGTGGCACTGTGTCTGGATCGA	1112
Qy	289	GluLysValProGluTyThrLeuThrIleGlnAlaThrAspMetAspGlyAsp-----	306
Db	1113	GAGAAAGTGCACAGTATACGTTAATAATTCAAGCCACAGACATGGAAGCAATCCCACCT	1172
Qy	307	---GlySerThrThrAlaValAlaValValGluIleLeuAspAlaAsnAspAsnAla	325
Db	1173	TATGGCCTTTCAAACACAGCCACAGCCGTATCATCGGTGACAGATGCAATGACAAATCCT	1232
Qy	326	ProMetPheAspProGlnLysTyrGluAlaHisValProGluAsnAlaValGlyHisGlu	345
Db	1233	CCAGAGTTTACTGCCATGACTTTCTACGGAGAAGTCCTCGAGAACAGGGTGGAGCTCATT	1292
Qy	346	ValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerProAlaTTrpArgAlaThr	365
Db	1293	GTACCCAACTTAAGTGTACAGGCAAAAGATCAGCCACAGCGCGCTCGGAATCGGCA	1352
Qy	366	TyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThrThrHisProGluSer	385
Db	1353	TACAGAACTCAGTGTGTGAGACCTTACAGAAAGTTTCCCATCTCTGACAGACCCCAACAGC	1412
Qy	386	AsnGlnGlyIleLeuThrThrArgGlyGlyLeuAspPheGluAlaTyrAsnGlnHisThr	405
Db	1413	AATGATGGGTAGTCACAGTGGTAAAAACCAATTGACTTTGAAACGAATAGGATGTTGTGTC	1472
Qy	406	LeuTyValGluValThrAsnGluAlaProPheValLeuLysLeu-----ProThr	422
Db	1473	CTTACTGTTGTCGAGAAACCAAGTGCCATTAGCTAAAGCAATCAGCACCCACCTCAG	1532
Qy	423	SerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPheValPro	442
Db	1533	TCGACAGCCACTGTGTCTGTGACAGTTATTGATGTCATGAAATCCTTATTTGCCCCA	1592
Qy	443	ProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCysValTy	462
Db	1593	AATCCTAAAAATCATTCGCCAAGGAAGGCGCTCCACGACGATCCATGCTGCACACGCTC	1652
Qy	463	ThrAlaGluAspProAspLys---GluAsnGlnLysIleSerTyArgIleLeuArgAsp	481
Db	1653	ACTGCTCAGACCCCGATCGATATATCGACAGNATATCAGATACACAAATTGTCTGAT	1712
Qy	482	ProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThrLeu	501
Db	1713	CCTGCCAACTGGCTGAAAATAGACCCCGTGAATGGGCAGATCACTACTATTGCGGTTTG	1772
Qy	502	AspArgGluAspGluGlnPheValArgAsnAsnIleTyThrGluValMetValLeuAlaMet	521
Db	1773	GACAGAAATCCGCAAT---GTAAAAACAACATCATTAATGCTACCTTCTTGTCTCT	1829
Qy	522	AspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAspVal	541
Db	1830	GACAATGGAATCCGCGCTATGAGTGGGACAGGAACACTGCAAAATCATTATTACTTGATATC	1869
Qy	542	AsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerProVal	561
Db	1890	AATGACAAACGCCCTCAGGTGTTACCTCAAGAGCGGAGACCTGT---GAAATCCAGAA	1946

Qy	562	ArgHisValLeuAsnIleThr	-----AspLysAspLeuSerProHisThrSerProPhe	579
Db	1947	CCCAACTCAATTAACATACACAGCATTGTATATGACATAGACCCAAACGCGGCCGCTTC	2006	
Qy	580	GlnAlaGlnLeu	-----ThrAspAspSerAspIleTyrTprThrAlaGlu	594
Db	2007	GCCTTTGATCTTCCCTTATCTCCAGTGACATTATAAAGAAAC	2054	
Qy	595	ValAsnGlu	---GluGlyAspThrValValLeuSerLeuLys---LysPheLeuLysGln	612
Db	2055	ATCAACCGGCTTAATGGTGATTTTGTCTCAGCTCAATTAAGATAAATAATTTTGGAAAGCT	2114	
Qy	613	AspThrTyrAspValHisLeuSerLeuSerAspHisGlyAsn	---LysGluGlnLeu	630
Db	2115	GGTATCTATGAGGTTCCCATCATATCACAGATTACGGGAATCCCCCAAGTCCAAACATT	2174	
Qy	631	ThrValIleArgAlaThrValCysAspCysHisGlyHis	-----ValGlu	645
Db	2175	TCCATCTCGGCTGTGAAGTTTGTGTCAGTGTGACTCCCAATGGAGACTGCACGGACGTGGAC	2234	
Qy	646	ThrCysProGlyProTrpLysGlyGlyPheIleLeuProValLeuGlyAlaValLeuAla	665	
Db	2235	AGGATCGTGGGTGCAGGGCTTGGC	2276	
Qy	666	LeuLeuPheLeuLeuValLeuLeuLeuVal	-----	677
Db	2277	ATCCTCTGTGTATCATCATCTCTGCTGATCCTTGTCTCATGTGTGGTATGGATGAAA	2336	
Qy	678	-----ArgLysLysArgLysIleLysGlyProLeuLeuLeuProGluLysAspThrArg	695	
Db	2337	CGCGGGGATTAAGAGCGCAAGCCAAAGCAGCATTATTAATCACCCAGAAGATGATGAAGA	2396	
Qy	696	AspAsnValPheTyrTyrGlyGluGluGlyGlyGluAspGlnAspTyrAspIle	715	
Db	2397	GATAATATATTGAAATATGATGAAGAAGGTGGAGAGAGAAGACCCAGGACTATGACTTG	2456	
Qy	716	ThrGlnLeuHisArgGlyLeuGluAlaArgProGluVal	-----ValLeu	730
Db	2457	AGGCAGCTCCAGCAACACAGATATCTGGAGCGCTGATGCCATCAAGCCCGTGGGAATCAGA	2516	
Qy	731	ArgAsnAspValAlaProThrIleIleProThrProMetTyrArgProArgProAla	---	749
Db	2517	CGGCTAGACAGAGGCT---ATCCATGTGAGCCACAGTACCAGTCCCGATCCGAGACC	2573	
Qy	750	---AsnProAspGluIleGlyAsnPheIleIleGluAsnLeuLysAlaAlaAsnThrAsp	768	
Db	2574	CCACACCTGGGATATTGGGACTTCATTAAATGAGGGCCTTAAAGCTGTGACAAACGAC	2633	
Qy	769	ProThrAlaProProTyrAspThrLeuLeuValPheAspTyrGluGlySerGlySerAsp	788	
Db	2634	CCACACGGCCACCGTATGATCCTCTTATGCTTTTACTACAGGCGCAGCGCTCCACG	2693	
Qy	789	AlaAlaSerLeuSerSerLeuThrSerSerAlaSerAspGlnAspTyrAspTyr	808	
Db	2694	GCTGGCTCCTTGAGCTCCCTCAACTCTCCAGTAGCGGTGGGACACAGGACTATGACTAC	2753	
Qy	809	LeuAsnGluTrpGlySerArgPheLysLysLeuAlaAspMetTyrGlyGlyGluAsp	828	
Db	2754	CTGAATGACTGGGGACCCCGCTTCAAGAAACTGGCGGACATGTACGCGCGTGTGACGAC	2813	
RESULT 6				
BM923882				
LOCUS				
DEFINITION				
AGENCOCURT_6709933 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5760016				
5', mRNA sequence.				
ACCESSION				
BM923882				
VERSION				
BM923882.1 GI:19374261				
KEYWORDS				
EST.				
SOURCE				
Homo sapiens (human)				
ORGANISM				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE				
1 (bases 1 to 1016)				

/note=Vector: pCMV-SPORT6; Site 1: EcoRV; Site 2: Not I;
Subtracted with brain, liver, lung, kidney and muscle.
Directionally cloned. Priming method: oligo-dr. Average
insert size: 1800 bp. Library amplification: 26,000 fold.
Kristi A. Eglund, James J. Vincent, Robert Strausberg,
Bungkook Lee & Ira Pastan: Discovery of new breast
cancer genes encoding membrane and secreted proteins.
Manuscript submitted."

ORIGIN

Alignment Scores:	4.88e-144	Length:	873
Pred. No.:	1459.00	Matches:	283
Score:	97.26%	Conservative:	1
Percent Similarity:	96.92%	Mismatches:	5
Best Local Similarity:	33.39%	Indels:	3
Query Match:	14	Gaps:	0
DB:			
US-09-916-849A-1 (1-829) x CA453911 (1-873)			
Qy	452	GlyLeuProThrGluProValCysValTyrThrAlaGluAspProAspLysGluAsn	471
Db	3	GGGATCCCACTGGGAGCCCTGTGTCTACACTGCAGAGACCCCTGCAGAGGAGAT	62
Qy	472	GlnLysIleSerTyrArgIleLeuArgAspProHlaGlyTrpLeuAlaMetAspProAsp	491
Db	63	CAAAAGATCATCTACCGCATCTCTGAGACCCAGCGGTGGCTAGCCATGGACCCAGAC	122
Qy	492	SerGlyGlnValThrAlaValGlyThrLeuAspArgGluAspGluGlnPheValArgAsn	511
Db	123	AGTGGCAGGTTCACAGCTGTGGGACCCCTGCACCGTGAGATGAGCACTTTGTGAGGAAC	182
Qy	512	AsnIleTyrGluValMetValLeuAlaMetAspAsnGlySerProProThrThrGlyThr	531
Db	183	AACATCTATGAGTCATGGTCTTGGCCATGGACAAATGGAAGCCCTCCACCACTGGCAGC	242
Qy	532	GlyThrLeuLeuLeuThrLeuIleAspValAsnAspHisGlyProValProGluProArg	551
Db	243	GGAAACCTTCTGTAACTGATGATGTCACGACCATGGCCAGTCCCTGAGCCCGCT	302
Qy	552	GlnIleThrLeuCysAsnGlnSerProValArgHisValLeuAsnIleThrAspLysAsp	571
Db	303	CAGATCACCATCTGCACCAAGAGCCCTGTGGCCAGGTGCTGAACATCAGGCAAGGAC	362
Qy	572	LeuSerProHisThrSerProPheGlnAlaGlnLeuThrAspAspSerAspIleTyrTrp	591
Db	363	CTGTCTCCCAACACTCCCTCTTCCAGGCCAGCTCACATGACTCAGACATCTACTGG	422
Qy	592	ThrAlaGluValAsnGluGluGlyAspThrValValLeuSerLeuLysLysPheLeuLys	611
Db	423	ACGGCAGAGGTCAACGAGGAGGTGACACAGTGGTCTGTCCCTGAGAGATTCCTGAAG	482
Qy	612	GlnAspThrTyrAspValHisLeuSerLeuSerAspHisGlyAsnLysGluGlnLeuThr	631
Db	483	CAGGATACATATGACGTGCACCTTCTCTGTCTGACCATGGCAACAAAGAGAGCTGACG	542
Qy	632	ValIleArgAlaThrValCysAspCysHisGlyHisValGluThrCysProGlyProTrp	651
Db	543	GTGATCAGGCCCACTGTGTGGACTGCGCATGGCCATGTCGAAACCTTCCCTGACCCCTG	602
Qy	652	LysGlyGlyPheLeuLeuProValLeuGlyAlaValLeuAlaLeuPheLeuLeuLeu	671
Db	603	AAAGGAGGTTCATCTCCCTCTGTGGGGGCTGTCTGGCTCTGTCTCTCTCTCTCTG	662
Qy	672	ValIleLeuLeuLeuValArgLysLysArgLysIleLysGluProLeuLeuLeuProGlu	691
Db	663	GTCTGTCTTTTGTGTGGAAGAGAGCGGAAGATCAAGAGGCCCTCTCTACTCCAGAA	722
Qy	692	AspAspThrArgAspAsnValPheTyrTyrGlyGlu-GluGlyGlyGlyGluAspGly	711
Db	723	GATGACACCCGTGACACGCTTCTACTATGGGGAAGAGGGGGTGGGAGAGACCA	782
Qy	711	nAspTyrAspIleThrGlnLeuHisArgGlyLeuGlu-AlaArgProGluValValLeu-	730

Db	783	GGACTATGATCACCACCACTCCACCAAGTGTGGAGCCCAAGCCCGAGGTTCTTCTCC	842
Qy	731	ArgAsnAspValAlaProThrIleLeuPro 740	
Db	843	CGCAATGAGTGGCAGCAACCAACATCATCCCG 872	
RESULT 8			
LOCUS	CA455046	887 bp	mRNA linear EST 12-NOV-2002
DEFINITION	AGENCOURT_10714428 MAPcL Homo sapiens cDNA clone IMAGE:6722552 5'; mRNA sequence.		
ACCESSION	CA455046		
VERSION	CA455046.1	GI:24905371	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	1 (bases 1 to 887)		
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .		
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs@mail.nih.gov Tissue Procurement: Kristi A. Eglund, Ira Pastan cDNA Library Preparation: Invitrogen Corp cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM14285 row: h column: 08 High quality sequence stop: 594.		

FEATURES

1..887
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6722552"
/cell_lines="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231,
hTERT-HME1, LNCap"
/lab_host="EMDH103"
/clone_lib="MAPcL"
/note=Vector: pCMV-SPORT6; Site 1: EcoRV; Site 2: Not I;
Subtracted with brain, liver, lung, kidney and muscle.
Directionally cloned. Priming method: oligo-dr. Average
insert size: 1800 bp. Library amplification: 26,000 fold.
Kristi A. Eglund, James J. Vincent, Robert Strausberg,
Bungkook Lee & Ira Pastan: Discovery of new breast
cancer genes encoding membrane and secreted proteins.
Manuscript submitted."

ORIGIN

Alignment Scores:	5.29e-142	Length:	887
Pred. No.:	1440.00	Matches:	284
Score:	96.97%	Conservative:	4
Percent Similarity:	95.62%	Mismatches:	5
Best Local Similarity:	32.96%	Indels:	4
Query Match:	14	Gaps:	0
DB:			
US-09-916-849A-1 (1-829) x CA455046 (1-887)			
Qy	287	AspArgGluLysValProGluTyrThrLeuThrIleGlnAlaThrAspMetAspGlyAsp	306
Db	3	GACCCGGAAAAAGTCCCTGAGTACACTGACCTCCAGGCCAGACATGATGGGAC	62
Qy	307	GlySerThrThrAlaValAlaValValGluIleLeuAspAlaAsnAspAlaPro	326
Db	63	GGCTCCACACCAACCGGAGTGGAGTAGTGGAGATCTTGTATGCCATGACATGCTCCC	122
Qy	327	MetPheAspProGlnLysTyrGluAlaHisValProGluAsnAlaValGlyHisGluVal	346

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

http://image.iimr.gov
plate: LLC2658 row: h column: 21
High quality sequence stop: 609.

FEATURES	SOURCE
1. Feature 1	Source 1
2. Feature 2	Source 2
3. Feature 3	Source 3
4. Feature 4	Source 4
5. Feature 5	Source 5
6. Feature 6	Source 6
7. Feature 7	Source 7
8. Feature 8	Source 8
9. Feature 9	Source 9
10. Feature 10	Source 10
11. Feature 11	Source 11
12. Feature 12	Source 12
13. Feature 13	Source 13
14. Feature 14	Source 14
15. Feature 15	Source 15
16. Feature 16	Source 16
17. Feature 17	Source 17
18. Feature 18	Source 18
19. Feature 19	Source 19
20. Feature 20	Source 20
21. Feature 21	Source 21
22. Feature 22	Source 22
23. Feature 23	Source 23
24. Feature 24	Source 24
25. Feature 25	Source 25
26. Feature 26	Source 26
27. Feature 27	Source 27
28. Feature 28	Source 28
29. Feature 29	Source 29
30. Feature 30	Source 30
31. Feature 31	Source 31
32. Feature 32	Source 32
33. Feature 33	Source 33
34. Feature 34	Source 34
35. Feature 35	Source 35
36. Feature 36	Source 36
37. Feature 37	Source 37
38. Feature 38	Source 38
39. Feature 39	Source 39
40. Feature 40	Source 40
41. Feature 41	Source 41
42. Feature 42	Source 42
43. Feature 43	Source 43
44. Feature 44	Source 44
45. Feature 45	Source 45
46. Feature 46	Source 46
47. Feature 47	Source 47
48. Feature 48	Source 48
49. Feature 49	Source 49
50. Feature 50	Source 50
51. Feature 51	Source 51
52. Feature 52	Source 52
53. Feature 53	Source 53
54. Feature 54	Source 54
55. Feature 55	Source 55
56. Feature 56	Source 56
57. Feature 57	Source 57
58. Feature 58	Source 58
59. Feature 59	Source 59
60. Feature 60	Source 60
61. Feature 61	Source 61
62. Feature 62	Source 62
63. Feature 63	Source 63
64. Feature 64	Source 64
65. Feature 65	Source 65
66. Feature 66	Source 66
67. Feature 67	Source 67
68. Feature 68	Source 68
69. Feature 69	Source 69
70. Feature 70	Source 70
71. Feature 71	Source 71
72. Feature 72	Source 72
73. Feature 73	Source 73
74. Feature 74	Source 74
75. Feature 75	Source 75
76. Feature 76	Source 76
77. Feature 77	Source 77
78. Feature 78	Source 78
79. Feature 79	Source 79
80. Feature 80	Source 80
81. Feature 81	Source 81
82. Feature 82	Source 82
83. Feature 83	Source 83
84. Feature 84	Source 84
85. Feature 85	Source 85
86. Feature 86	Source 86
87. Feature 87	Source 87
88. Feature 88	Source 88
89. Feature 89	Source 89
90. Feature 90	Source 90
91. Feature 91	Source 91
92. Feature 92	Source 92
93. Feature 93	Source 93
94. Feature 94	Source 94
95. Feature 95	Source 95
96. Feature 96	Source 96
97. Feature 97	Source 97
98. Feature 98	Source 98
99. Feature 99	Source 99
100. Feature 100	Source 100

```

1. . . . .
/organism="Homo sapiens"

```

```
/organelle="mitochondrion"
/mol_type="tRNA"
```

```
/db xref="taxon-9606"
```

```
/opt_xref="C:\XREF"
/opt_xref="IMAGE.C:\XREF"
```

/clone="IMAGE:64/9/80"

```

/tissue_type="carcinoma, cell_line"

```

```
/lab_host="DH10B (phage-
```

```
/clone_lib="NIH_MGC_40"
```

/note="Organ; Vector: pOTB7; Site.1: XhoI;
Site.2: EcoRI; cDNA made by oligo-*dt* priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adapter: GGACGACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using Zap-CDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH MGC Library."

ORIGIN

Alignment Scores:		
Pred. No.:	1 75e-139	Length: 1049
Score:	141: 50	Matches: 292
Percent Similarity:	89.40%	Conservative: 7
Best Local Similarity:	89.30%	Mismatches: 28
Query Match:	32.44%	Indels: 6
DB:	13	Gaps: 0

UIS-09-916-849A-1 (1-829) X B0936686 (1-1049)

452	Qy	GlyIleProThrGlyGluProValCysValTyrThrAlaGluAspProAspLysGluAsn	471
3	Db	GGCATCCCCACTGGGAGCGCTGTGTGTCTACACTGCAGAACCCCTGCACAGGAGAAT	62
472	Qy	GlnIyAlIeSerTyrArgIleLeuArgAspProAlaGlyTrpLeuAlaMetAspProAsp	491
63	Db	CAAAAGATCAGCTACCGCATCTCCAGAGAGACCCAGCAGGGTGGTAGCCATGAGCCACGAC	122
492	Qy	SerGlyGlnValThrAlaValGlyThrLeuAspArgGluAspGluGlnPheValArgAsn	511
123	Db	AGTGGGCAAGTCACGCTGTGGGCACCTCGACCGTGAGGATGAGCAGTTTGTGAGGAAC	182
512	Qy	AsnIleTyrGluValMetValLeuAlaMetAspAsnGlySerProProThrThrGlyThr	531
183	Db	AACATCTATGAAGTCATGGTCTTGGCCATGGACAATGGAAGCCCTCCACCATCTGGCAGC	242
532	Qy	GlyThrLeuLeuLeuThrLeuIleAspValAsnAspHisGlyIleProValProGluProArg	551
243	Db	GGAAACCCCTCTGCTAAACACTGATTGATGTCAATGACCATGGCCCGAGTCCCTGAGCCCCGT	302
552	Qy	GlnIleThrIleCysAsnGlnSerProValArgHisValLeuAsnIleThrAspLysAsp	571
303	Db	CAGATCACCATCTGCACACCAAGCCCTGTGGCCAGGTGTGAACATCATCGGACAAGGAC	362
572	Qy	LeuSerProHisThrSerProPheGlnAlaGlnLeuThrAspAspSerAspIleTyrTrp	591
363	Db	CTGTCTCCCCACACCTCCCTCTTCCAGGCCCGAGCTCAGATGACTCAGACATCTACTGG	422
592	Qy	ThrAlaGluValAsnGluGlyAspThrValValLeuSerLeuLysLysPheLeuLys	611
423	Db	ACGGCAGAGGTCAACGAGGAAGGTGCACAGTGGTCTTGTCCCTGGAAGAAGTTCCTGAAG	482
612	Qy	GlnAspThrTyrAspValHisLeuSerLeuSerAspHisGlyAsnLysGluGlnLeuThr	631
483	Db	CAGATACATATGACGTGCACCTTCTCTGTCTGACCATGGCAACAAGAGCAGCTCAGC	542
632	Qy	ValIleArgAlaThrValCysAspCysHisGlyHisValGluThrCysProGlyProTrp	651
543	Db	GTGATCAGGGCCACTGTGTGGCATGGCATGGCATGCGCATGCGAAACCTCCCTGGACCTCGG	602

652 LysGlyGlyPheIleLeuProValLeuGlyAlaValLeuAlaLeuLeuPheLeuLeu 671
603 AAGGAGTTTCATCTCCCTGCTGGGGGCTGCTGCTGCTGCTGCTGCTGCTG 662
672 ValLeuLeuLeuLeuValArgLysLysArgLysLysLysLysLysLysLys 691
663 GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
692 AspAspThrArgAspAsnValPheTyrTyrGlyGluGluGlyGlyGlyGln 711
721 AT-GACACCCCGTGACACGCTCTTCTACTATGCGGAGAGGGGGTGGCGAAGACAG- 778
712 AspTyrAspIleThrGlnLeuHisArgGlyLeuGluAlaArgProGluValValLeuArg 731
779 GACTATGACATCCG-CAGCTTCCAGGAGTCTGAGAGCCAGGCGGAGTGTCTCCGC 837
732 AsnAspValAlaProThrIleLeuProThrProMetTyrArgProArgProAlaAsnPro 751
838 AATGACGTGA-CAACCATATATCCGGAACACCATGACCGCTCTTCCGCCACCAACCCA 896
752 AspGluIleGlyAsnPheIleLeuGluAsnLeuLysAlaAlaAsnThrAspProThrAla 771
897 GATGAAACGCCACCTTTTATATTGAAACCTGAAGGCGTTAAAAAGACCCCAAGGCC 956
772 ProProTyrAspThrLeuLeu 778
957 CGGCCCTAAGAACACCTTCTT 977

RESULT 10
BQ927154
LOCUS
DEFINITION BQ927154 912 bp mRNA linear EST 20-AUG-2002
AGENCOURT_8776344 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6382032
5', mRNA sequence.
BQ927154
VERSION BQ927154.1 GI:223342185
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabes-remail.nih.gov
Tissue Procurement: DCTD/DTF
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LICM2573 row: d column: 01
High quality sequence start: 283
High quality sequence stop: 659.
Location/Qualifiers
1..912
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6382032"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC 40"
note="Organ: prostate; Vector: pOTB7; Site 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).

FEATURES

source

Note: this is a NIH_MGC Library."

ORIGIN
Alignment Scores:
Pred. No.: 1,41e-138 Length: 912
Score: 1408.00 Matches: 265
Percent Similarity: 96.39% Conservative: 2
Best Local Similarity: 95.67% Mismatches: 10
Query Match: 32.23% Indels: 0
DB: 13 Gaps: 0
US-09-916-849A-1 (1-829) x BQ927154 (1-912)
QY 318 lleLeuAspAlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisVal 337
Db 49 ATCCCTTGATGCAATGACAAATGCTCCCATGTTGACCCCAAGAGTAGAGGCCCATGTG 108
QY 338 ProGluAsnAlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaPro 357
Db 109 CTTGAGATGAGTGGGCGCATGAGGTGACAGGCTGACGGTCACTGATCTGGAGCCCCC 168
QY 358 AsnSerProAlaTyrArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPhe 377
Db 169 AACTCACCAGCGTGGCGTCCACCTTATCATGGCGGTGACGAGCGGGACCATTTT 228
QY 378 ThrIleThrThrHisProGlySerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAsp 397
Db 229 ACCATCACCCACCCCTGAGAGCAACAGGCGATCTGACCAACCAAGAGGTTTGGAT 288
QY 398 PheGluAlaLysAsnGlnHisThrLeuTyrValGluValThrValThrAsnGluAlaProPheVal 417
Db 289 TTGTGTTTCAAAACACAGCACACCTGTACGTTGAAGTGACCAACGAGGCCCTTTTGTG 348
QY 418 LeuLysLeuProThrSerThrAlaThrIleValHisValGluAspValAsnGluAla 437
Db 349 CTGAAGCTCCCAACCTCCACAGCCACCATAGTGGTCCACGTGGAGGATGTGAATGAGGCA 408
QY 438 ProValPheValProProSerLysValValGluValGlnGluGlyIleProThrGlyGlu 457
Db 409 CTGTGTTTGTCCACCTCCAAAGTCGTTGAGTCCAGGAGGGCATCCCCACATGGGAG 468
QY 458 ProValCysValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArg 477
Db 469 CTTGTGTGTCTACACTGCAGAGACCCCTGACAAAGGAGAATCAAAAGATCAGCTACCGC 528
QY 478 lleLeuArgAspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAla 497
Db 529 ATCTGAGAGACCCAGCAGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 588
QY 498 ValGlyThrLeuAspArgGluAspGluInPheValArgAsnAsnIleTyrGluValMet 517
Db 589 GTGGGCACCCCTCGACCGTGGAGATGAGCAGTTTGTGAGGAAACAACATCTATGAAATCATG 648
QY 518 ValLeuAlaMetAspAsnGlySerProThrThrThrGlyThrGlyThrLeuLeuLeuThr 537
Db 649 GTCTTGGCCATGGAATGGAAGCCCTCCCACTGGACCGGAAACCTCTCTGTTAACA 708
QY 538 LeuIleAspValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsn 557
Db 709 CTGATTGATGTCATGACCATGCGCCAGTCCCTGAGCCCGTCCAGATCACCATCTGCAAC 768
QY 558 GlnSerProValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSer 577
Db 769 CAAGCCCTGTGCGCGAGTGTGAACATCAAGGCAAGGACCTGTCTCTCCCAACCTCC 828
QY 578 ProPheGlnAlaGlnLeuThrAspSerAspIleTyrTyrThrAlaGlu 594
Db 829 CTTTTCAGGGCCCGAGCTCCAGATGACTGACATCTACTGGAACGGAAAA 879

RESULT 11
BQ690885
LOCUS
DEFINITION BQ690885 846 bp mRNA linear EST 15-JUL-2002
AGENCOURT_8064585 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6206222

Subtracted with brain, liver, lung, kidney and muscle.
 Directionally cloned. Priming method: oligo-dt. Average
 insert size: 1800 bp. Library amplification: 26,000 fold.
 Kristi A. Egland, James J. Vincent, Robert Strausberg,
 Bungkook Lee & Ira Pastan: Discovery of new breast
 cancer genes encoding membrane and secreted proteins.
 Manuscript submitted."

ORIGIN

Alignment Scores:

Pred. No.: 7,42e-136 Length: 918
 Score: 1382.50 Matches: 284
 Percent Similarity: 95.03% Conservative: 3
 Best Local Similarity: 94.04% Mismatches: 7
 Query Match: 31.84% Indels: 8
 DB: 14 Gaps: 3

US-09-916-849A-1 (1-829) x CA454560 (1-918)

Qy 180 LysTyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMet 139
 Db 3 AAGTATGAGTCTTTGGCCACGCTGTGTACAGATGTCCTCAGTGGAGGCCCATG 62
 Qy 200 AsnIleSerIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThr 219
 Db 63 AACATCTCCATCATCGTGCACGACCAATGACCAAGCCCAAGTTTACCCAGGACAC 122
 Qy 220 PheArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAla 239
 Db 123 TTCGAGGGAGTGCTTTAGAGGGAGTCTCTACAGGTACTTCTGTGATGAGGTGACAGCC 182
 Qy 240 ThrAspGluAspAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSer 259
 Db 183 ACGATGAGGATGATGCCATCTACACCTACATGAGGTGTGTCTACTCATCCATAGC 242
 Qy 260 GlnGluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIle 279
 Db 243 CAAGAACCAAGGACCCACACGACCTCATGTTCACCATTCACCGGAGCACAGGCACATC 302
 Qy 280 SerValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGln 299
 Db 303 AGCTGTATCTCCAGTGCCTGGACGGGAAAGTCCCTCAGTACACACGTGACATCCAG 362
 Qy 300 AlaThrAspMetAspGlyAspGlySerThrThrAlaValAlaValAlaValGluIleLeu 319
 Db 363 GCCACACATGATGGGACGGCTCCACACACGCGAGTGGCAGTAGTGGAGATCCTT 422
 Qy 320 AspAlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGlu 339
 Db 423 GATGCCAATGACAATGCTCCCATGTTTGACCCCCAGAAAGTACGAGGCCCATGTGCTGAG 482
 Qy 340 AsnAlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSer 359
 Db 483 AATGACATGGGCCATGAGGTGCAGAGGCTGACGTCACTGATCTGGACGCCCCCACTCA 542
 Qy 360 ProAlaTrpArgAlaThrTyrLeuIleMetGlyAspAspGlyAspHisPheThrIle 379
 Db 543 CCAGCGTGGCGTGCACCTACCTTATCATGGCGGTGACGACGGGACCATTTTACCATC 602
 Qy 380 ThrThrHisProGluSerAsnGlnGlyIleLeuThrThrArg-LysGlyLeuAspPheG1 399
 Db 603 ACCACCCACCTGAGACCAACAGGCGATCTGACACACGAGGAGGTTTGGATTGTA 662
 Qy 399 uAlaLysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeu 419
 Db 663 GGCCAAAACACGACACCTGTACGTGAAGTGACCAACGAGGCCCTTTTGTGTGAA 722
 Qy 419 sLeuProThrSerThrAlaThrIleValValHis-ValGluAspValAsnGluAlaPro 439
 Db 723 GCTCCCAACCTCCACAGCCACCATAGTGGTCCAGGTGGAGGATGTGAATGGAGCACTG 782
 Qy 439 alPheValPro-ProSerLysValVal-GluValGln-GluGlyIlePro---ThrGlyG 457

Db 783 TGTGTGTCCACCCCTCCANAGTCGTGTGAGGTCGAGGAGGGCATTCCTCCCTCGGGGA 842
 Qy 457 luProVal---CysValTyrThrAlaGluAspPro---AspLysGluAsnGlnLys 473
 Db 843 AGCTGTGTGTGTCTACACCTGGCAGAGACCCCTGGCAAGGAAGATTCAAA 858

RESULT 13

BQ928283

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 912)

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs-remail.nih.gov

Tissue Procurement: DCTD/DTF

cDNA Library Preparation: Rubin Laboratory

DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCM2663 row: d column: 08

High quality sequence stop: 585.

Location/Qualifiers

1..912

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:6481591"

/tissue_type="carcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH_MGC_40"

/note="Organ: prostate; Vector: pOTB7; Site: 1: XhoI;

Site 2: EcoRI; cDNA made by oligo-dt priming.

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGCACGAG(G). Library constructed by

Ling Hong in the laboratory of Gerald M. Rubin (University

of California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies).

Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:

Pred. No.: 1,53e-135 Length: 912
 Score: 1379.50 Matches: 282
 Percent Similarity: 93.11% Conservative: 2
 Best Local Similarity: 92.46% Mismatches: 11
 Query Match: 31.57% Indels: 10
 DB: 13 Gaps: 2

US-09-916-849A-1 (1-829) x BQ928283 (1-912)

Qy 452 GlyIleProThrGlyGluProValCysValTyrThrAlaGluAspProAspLysGluAsn 471
 Db 3 GGCATCCCCACCTGGGGAGCCTGTGTGTCTACCTGCAGAGACCCCTGACAAGAGAT 62
 Qy 472 GlnLysIleSerTyrArgIleLeuArgAspProAlaGlyTyrLeuAlaMetAspProAsp 491
 Db 63 CAAAGATCAGTACCGCATCTCTGAGACCCAGAGGGTGGCTAGCATGACCCAGAC 122
 Qy 492 SerGlyGlnValThrAlaValGlyThrLeuAspArgGluAspGluGlnPheValArgAsn 511
 Db 123 AGTGGGAGGTACACAGTGTGGCCACCTCGACCGTGGAGTACGACGTTTGTGAGGAAC 182

QY 512 AsnIleTyrGluValMetValLeuAlaMetAspAsnGlySerProThrThrGlyThr 531
 Db 183 AACATCATGAAGTCATGGTCTTGGCCATGGACATGAAGCCCTCCACCACCTGGCAGC 242
 QY 532 GlyThrLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 551
 Db 243 GGAACCCCTTCGTAACACTGATGATGATCAATGATGATGATGATGATGATGATGAT 302
 QY 552 GlnIleThrIleCysAsnGlnSerProValArgHisValLeuAsnIleThrAspLysAsp 571
 Db 303 CAGATCACCATCTGCAACCAAGCCCTGTGGCCAGGTGTGACATCAGGACAGGAC 362
 QY 572 LeuSerProHisThrSerProPheGlnAlaGlnLeuThrAspSerAspIleTyrTyr 591
 Db 363 CTGTCTCCCCACAGCTCCCTTTCCAGGCCAGGTCTCAGATGACTCAGACATCTACTGG 422
 QY 592 ThrAlaGluValAsnGluGluValAspThrValValLeuSerLeuLysLysPheLeuLys 611
 Db 423 ACGCAGAGGTTCACGAGGAAGTGCACAGTGTCTTGTCCCTGAGAAAGTTCCTGAG 482
 QY 612 GlnAspThrTyrAspValHisLeuSerLeuSerAspHisGlyAsnLysGluGlnLeuThr 631
 Db 483 CAGGATACATATGACGTGCACCTTCTGTCTGACCATGGCAACAAAGACGCTGAGC 542
 QY 632 ValIleArgAlaThrValCysAspCysHisGlyHisValGluThrCysProGlyProTyr 651
 Db 543 GTGATCAGGGCCACTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 602
 QY 652 LysGlyGlyPheIleLeuProValLeuGlyValValLeuAlaLeuLeuPheLeuLeu 671
 Db 603 AAAGGAGGTTTCATCTCCCTGCTGCTGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 662
 QY 672 ValLeuLeuLeuLeuValArgLysArgLysLysLysLysLysLysLysLysLysLys 691
 Db 663 GTGCTGCTTTTGTGGTGAGAAAAAGCGGAAGATCAAGGAGGCCCTCTACTCCCGAA 722
 QY 692 AspAspThr-ArgAspAsnValPheTyrTyrGlyGluGluGlyGlyGlyGlyGlyGly 711
 Db 723 GATGACACCCGTCGACACGCTCTTACTATGCGGAAAGGGGGTGGCGGAAAGGACC 782
 QY 711 In-AspTyrAspIleThr-GlnLeuHisArg-GlyLeuGlu-AlaArgPro---GluVal 728
 Db 783 AGGACTATGACATCACCCAGCTCCACCGAGGTCTGGAAGGCCAGGCGCGGAAAGTGGG 842
 QY 729 ValLeuArg-AsnAspValAla-----ProThrIleIleProThrProMetTyrArgPr 746
 Db 843 TTCCTCCGAAATGACGTGGGACCAACCAACCATCTTCCCGAAACCATGGTACGGTCC 902
 QY 746 o 746
 Db 903 C 903
 RESULT 14
 CA489015
 LOCUS
 DEFINITION CA489015 875 bp mRNA linear EST 14-NOV-2002
 AGENCOURT_10809410 MAPCL Homo sapiens cDNA clone IMAGE:6721249 5',
 mRNA sequence.
 CA489015
 VERSION
 CA489015.1 GI:24951806
 KEYWORDS
 EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 875)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 JOURNAL
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgaabbs-remail.nih.gov
 Tissue Procurement: Kristi A. Eglund, Ira Pastan
 CDNA Library Preparation: Invitrogen Corp

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM14282 row: b column: 01
 High quality sequence stop: 647.
 Location/Qualifiers
 1. 875
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6721249"
 /cell_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231,
 hTERT-HWE1, LNCap"
 /lab_host="EMDH10B"
 /clone_lib="MAPCL"
 /note="Vector: pCMV-SPORT6; Site 1: EcoRV; Site 2: Not I;
 Subtracted with brain, liver, lung, kidney and muscle.
 Directionally cloned. Priming method: oligo-dT. Average
 insert size: 1800 bp. Library amplification: 36,000 fold.
 Kristi A. Eglund, James J. Vincent, Robert Strausberg,
 Bungkook Lee & Ira Pastan: Discovery of new breast
 cancer genes encoding membrane and secreted proteins.
 Manuscript submitted."

ORIGIN

Alignment Scores:
 Pred. No.: 1,978-132 Length: 875
 Score: 1350.00 Matches: 262
 Percent Similarity: 98.87% Conservativity: 1
 Best Local Similarity: 98.50% Mismatches: 2
 Query Match: 30,90% Indels: 2
 DB: 14 Gaps: 0
 US-09-916-849A-1 (1-829) x CA489015 (1-875)
 QY 565 LeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGlnAlaGluLeuThr 584
 Db 3 CTGAACATCAGGACAGGACCTGTCTCTCCCC-ACCTCCCTTTCCAGGCCAGCTCACA 61
 QY 585 AspAspSerAspIleTyrTyrThrAlaGluValAsnGluGlyAspThrValValLeu 604
 Db 62 GATGACTCAGACATCTACTGACGCGGAGGTCAACAGGAGGAGGTGACACAGTGTCTTG 121
 QY 605 SerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSerLeuSerAspHis 624
 Db 122 TCCTTGAAGAGTTCCTCGACGAGATACATATGACGTGACCTTCTCTCTGACCAT 181
 QY 625 GlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCysHisGlyHisVal 644
 Db 182 GGCAACAAAGAGCAGCTGACGTGATCAGGGCCACTGTGTGGACTGCGCATGGCCATGTC 241
 QY 645 GluThrCysProGlyProTyrLysGlyGlyPheIleLeuProValLeuGlyValValLeu 664
 Db 242 GAAACCTGCTGACCTCTGAGAGGAGGTTCATCTCTCTCTCTCTCTCTCTCTCTCTG 301
 QY 665 AlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysLysArgLysLysLys 684
 Db 302 GCTCTGCTGTCT 361
 QY 685 GluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyrTyrGlyGlu 704
 Db 362 GAGCCCT 421
 QY 705 GlyGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisArgGlyLeuGluAla 724
 Db 422 GGGGGTGGGAGAGGAGCAGGACTATGACATCACAGCTCCAGCTCCAGCTCTCTCTCTCT 481
 QY 725 ArgProGluValValLeuArgAsnAspValAlaProThrIleIleProThrProMetTyr 744
 Db 482 AGGCCGAGGTGTCT 541

```

QY 745 ArgProArgProAlaAsnProAspGluIleGlyAsnPhelIleGluAsnLeuLysAla 764
Db 542 CGTCTCGGCGAGCCCAACAGATGAAATCGCACTTTATTAATGAGAACTGAGGCG 601
QY 765 AlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPheAspTyrGluGly 784
Db 602 GCTAACAGACCCCAAGCCCGCCCTTACGACACCCCTCTGGTGTTCGATATGAGGCG 661
QY 785 SerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSerAspGln 804
Db 662 AGCGGCTCCGAGCGCGCTGCTGAGCTCCCTCACTCTCTCGGCTCCGACAGACCCA 721
QY 805 AspTyrAspTyrLeuAsnGluTyrGlySerArgPheLysLysLeuAlaAspMetTyr-GI 824
Db 722 GATTACGATTATCTGAACAGTGGGCGAGCCGCTTCAAGAACTGGCAGACATGTACCG 781
QY 824 YGlyGlyGluAspAsp 829
Db 782 TGGCGGGGAAGAGAC 797

RESULT 15
BQ676796
LOCUS
DEFINITION BQ676796 913 bp mRNA linear EST 15-JUL-2002
5', mRNA sequence.
AGENCOURT 8184299 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6264044
BQ676796
VERSION BQ676796.1 GI:21789475
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 913)
NH-MGC http://mgs.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DTF
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM2428 row: o column: 21
High quality sequence stop: 635.
FEATURES
source
1..913
location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6264044"
/tissue_type="melanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_112"
/note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCACGAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
ORIGIN
Alignment Scores:
Pred. No.: 1.48e-131 Length: 913
Score: 1343.00 Matches: 277
Percent Similarity: 94.93% Conservative: 4
Best Local Similarity: 93.58% Mismatches: 6
Query Match: 30.74% Indels: 9
DB: 13 Gaps: 2

```

Search completed: September 23, 2004, 01:24:50
Job time : 5141 secs

US-09-916-849A-1 (1-829) x BQ676796 (1-913)

```

QY 275 SerThrGlyThrIleSerValIleSerSerGlyLeuAspArgGluLysValProGluTyr 294
Db 2 AGCACAGGACCATCAGCGTCATCTCCAGTGGCTGGACCGGAAAGTCCCTGAGTAC 61
QY 295 ThrLeuThrIleGlnAlaThrAspMetAspGlyAspGlySerThrThrAlaValAla 314
Db 62 ACACCTGACCATCCAGGCGCACAGCATGGATGGGAGCGGTCCACACACGACGAGTGGCA 121
QY 315 ValValGluIleLeuAspAlaAsnAspAlaProMetPheAspProGlnLysTyrGlu 334
Db 122 GTAGTGGAGATCCTTATGATGCAATGCAATGCTCCCATGTTTACCCCCAGAGTACGAG 181
QY 335 AlaHisValProGluAsnAlaValGlyHisGluValGlnArgLeuThrValThrAspLeu 354
Db 182 GCCCATGTGCTGAGAAATGAGTGGGCCATGAGGTGCAGAGGTGACGGTCACTGATCTG 241
QY 355 AspAlaProAsnSerProAlaTyrArgAlaThrTyrLeuIleMetGlyGlyAspAspGly 374
Db 242 GACGCCCCCAACTCACAGCGTGGCGTGCACCTACCTTATCATGGCGGTGACACGCG 301
QY 375 AspHisPheThrIleThrThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLys 394
Db 302 GACCATTTTACCATCACCCACCCCTGAGAGCAACCCAGGGCATCTGTGACACACCAAGAA 361
QY 395 GlyLeuAspPheGluAlaLysAsnGlnHisThrLeuTyrValGluValThrAsnGluAla 414
Db 362 GGTTCGATTTTGGAGCCCAAAACCCAGCACACCTGTAGTTGAAGTGACCAACGAGGCC 421
QY 415 ProPheValLeuLysLeuProThrSerThrAlaThrIleValValHisValGluAspVal 434
Db 422 CTTTTTGTCTGAAGCTCCCAACCTCCACAGCACCATAGTGTGTCACACGTGGAGGATGTG 481
QY 435 AsnGluAlaProValPheValProProSerLysValValGluValGlnGluGlyIlePro 454
Db 482 ATGAGGCACCTGTGTTGTGCCACCTCCAAAGTCGTTGAGTCCAGGAGGCGATCCCC 541
QY 455 ThrGlyGluProValCysValTyrThrAlaGluAspProAspLysGluAsnGlnLysIle 474
Db 542 ACTGGGGAGCCTGTGTGTCTTACACTGCAGAAAGACCCCTGCACAGGAGAAATCAAAAGATC 601
QY 475 SerTyrArgIleLeuArgAspProAlaGlyTyrLeuAlaMetAspAspSerGlyGln 494
Db 602 AGTACCGCATCTCTGAGAGACCCAGAGGTGGCTAGCCATGACCCACGACAGTGGGCGAG 661
QY 495 ValThrAlaValGlyThrLeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyr 514
Db 662 GTCACAGCTGTGGGCACCTCGACCGTGAGGATGAGCAGTTTGTGAGGAAACACATCAT 721
QY 515 GluValMetValLeuAlaMetAspAsnGlySerProProThrThrThrGlyThrGlyThrL 534
Db 722 GAAGTCATGGCCTTGTCCCATGGACATGAAAGCCCTCCACCCCTCCGACGCGGAAACCC 781
QY 534 euleLeuThrLeu-IleAspVal-AsnAsp-HisGly-ProValPro---GluProArg 551
Db 782 TTCTGCTAACACTGATGATGATGTCATGACCCATGGCCCCCATCCCTGGAGGCCCCCGT 841
QY 552 Gln-IleThr-----IleCysAsnGlnSerProVal 561
Db 842 CAGGATCATCCCATCTTGTGCCAACAAAGCCCGCTG 875

```

This Page blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 22, 2004, 20:44:18 ; Search time 8138 Seconds
(without alignments)
4415.259 Million cell updates/sec

Title: US-09-916-849A-1
Perfect score: 4369
Sequence: 1 MGPRGPFLASILLQVCWLQ.....NEWGSRFKKLADMYGGGEDD 829

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool_h/US09916849/runat_21092004_145803_7009/app_query.fasta_1.967
-DB=GenEmbl -OFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORV=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09916849@cgn_1_1_5239@runat_21092004_145803_7009 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.*
1: gb_ba.*
2: gb_hg.*
3: gb_in.*
4: gb_em.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*

29: em_vi.*
30: em_hig_hum.*
31: em_hig_inv.*
32: em_hig_other.*
33: em_hig_mus.*
34: em_hig_pln.*
35: em_hig_rod.*
36: em_hig_mam.*
37: em_hig_vrt.*
38: em_sy.*
39: em_hgo_hum.*
40: em_hgo_mus.*
41: em_hgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	4369	100.0	3171	6	AR270785	Sequence
2	4369	100.0	3171	6	AX411115	Sequence
3	4369	100.0	3171	6	AX829232	Sequence
4	4369	100.0	3171	9	HSPCAD	X63629 H.sapiens m
5	4358	99.7	3686	9	BC041846	BC041846 Homo sapi
6	4348	99.5	3219	6	AX382118	Sequence
7	3550.5	81.3	3187	10	WMCADHP	X06340 Mouse mRNA
8	3433.5	78.6	2993	10	BC052189	Mus muscu
9	2549.5	58.4	2436	5	GGCADHB	X58518 Chicken mRN
10	2420	55.4	4778	6	AX658364	Sequence
11	2420	55.4	4778	9	HSECAD	Z13009 H.sapiens m
12	2420	55.4	4828	6	AX587858	Sequence
13	2417	55.3	2649	9	HSECADH	Z18923 H.sapiens m
14	2400.5	54.9	2807	9	AB025106	Sequence
15	2388	54.7	4362	10	YMECADH	AB025106 Homo sapi
16	2388	54.7	4362	10	AB017696	X06115 Mouse mRNA
17	2377	54.4	3450	5	XL004707	AB017696 Rattus no
18	2375	54.4	2699	5	XL004707	U04707 Xenopus lae
19	2371	54.3	3750	5	XL004707	X63720 X.laevius mR
20	2349.5	53.8	2808	6	IS7999	X78546 X.laevius XB
21	2349.5	53.8	2808	6	AR214053	Sequence 1
22	2349.5	53.8	2808	6	AR369494	Sequence
23	2342.5	53.6	2808	6	HUMVORCAD	AR369494 Sequence
24	2323	53.2	2106	4	BTCDHP	L08599 Human uvom
25	2274	52.0	3545	5	CHXLCAMR	X53614 Bovine mR
26	2223	50.9	2486	10	MMUVOM	M16260 Chicken liv
27	2216.5	50.7	2567	9	AB025105	X06339 Mouse mRNA
28	2204.5	50.5	4097	5	XELCADH	AB025105 Homo sapi
29	2198	50.3	2819	5	XL004708	L29057 Xenopus lae
30	2142	49.0	2173	5	XLNRECAD	U04708 Xenopus lae
31	2131	48.8	4434	5	CHKKCAM	X75454 X.laevius mR
32	1893	43.3	3340	5	AF364811	M81894 Chicken B-c
33	1707	39.1	2853	9	HSNCAD	AF364811 Danio rer
34	1704.5	39.0	3203	5	GGNCAD	X54315 Human mRNA
35	1702	39.0	4033	9	BC036470	X07277 Chicken mRN
36	1701.5	38.9	4350	10	AB017695	BC036470 Homo sapi
37	1701	38.9	3269	10	BC022107	AB017695 Rattus no
38	1701	38.9	4321	10	AB008811	BC022107 Mus muscu
39	1699.5	38.9	4132	9	S42303	AB008811 Mus muscu
40	1697.5	38.9	4321	6	AX305854	S42303 N-cadherin
41	1697.5	38.9	4321	10	MUSCADNA	AX305854 Sequence
42	1694	38.8	2954	5	CHKRCG	M31131 Mouse neur
43	1693	38.8	3036	10	AF097593	D14459 Gallus gall
44	1691	38.7	3839	4	BTCDHP	AF097593 Rattus no
45	1682	38.5	2827	5	AF418565	X53615 B.taurus mR
						AF418565 Danio rer

ALIGNMENTS

RESULT 1


```

Db 1794 GCCAGCTCAGAGATGACTCAGATCTACTGCGCGGAGAGGTCAACGAGAGGTGAC 1853
QY ThrValValLeuSerLeuLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
Db 1854 ACAGTGTCTTGTCCCTGAGAGAGTTCTCTGAGCAGAGTATATAGAGTGCACCTTCT 1913
QY LeuSerAspHisGlyAsnLysGlnLeuThrValLysArgAlaThrValCysAspCys 640
Db 1914 CTGTCTGACCATGGCAACAAAGAGCAGCTGACGGTGTATCAGGGCCTGTGTGGAGTGC 1973
QY HisGlyHisValGluThrCysProGlyProTyrPheGlyGlyPheLeuLeuProValLeu 660
Db 1974 CATGCCATGTGCAAACTGCTGACCTTGGACCTTGGAAAGAGTTTCACTCCCTGTGCTG 2033
QY GlyAlaValLeuAlaLeuLeuPheLeuLeuLeuValLeuLeuLeuValArgLysLys 680
Db 2034 GGGGTGTGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2093
QY ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
Db 2094 CGGAAGATCAAGGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2153
QY TyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
Db 2154 TATGGCAAGAGGGGGTGGGGAAGAGACCCAGACTATGACATCACCCAGCTCCACCGA 2213
QY GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleLeuPro 740
Db 2214 GGCTCGAGGCGGCGGAGGGTGGTCTCCGCAATGACGTGGCACCACCATCATCCCG 2273
QY ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleGlu 760
Db 2274 ACACCATGTACCTCTCTAGCCAGCCACCCAGATGAAATCGGCACTTTAATTTAG 2333
QY AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
Db 2334 AACCTGAAGCGGGTAAACACAGACCCACAGCCCGCCCTAGCACCCCTCTTGGTCTC 2393
QY AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
Db 2394 GACTATGAGGACCGGCTCCGACCGCGGCTCTGAGCTCCCTGAGCTCCCTCTCCGCTCC 2453
QY AspGlnAspGlnAspTyrAspTyrLeuAsnGluTyrGlySerArgPheLysLysLeuAla 820
Db 2454 GACCAAGACCAAGATTACGATTATCTGACAGTGGGGGAGCCGCTTCAGAGAGCTGGCA 2513
QY AspMetTyrGlyGlyGlyAspAsp 829
Db 2514 GACATGTACGTGGCGGGAGGAGCAC 2540

RESULT 2
AX411115 3171 bp DNA linear PAT 14-JUN-2002
LOCUS
DEFINITION Sequence 3762 from Patent WO0229103.
ACCESSION AX411115
VERSION AX411115.1 GI:21443820
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.
TITLE Gene expression profiles in liver cancer
JOURNAL Patent: WO 0229103-A 3762 11-APR-2002;
GENE LOGIC INC (US)
FEATURES
source
1. .3171
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon.9606"
/note="EMBL/GenBank Accession No. X63629"

```

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 3171
 Score: 4369.00 Matches: 829
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-09-916-849A-1 (1-829) x AX411115 (1-3171)

QY 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuLeuValCysTrpLeuGln 20
 Db 54 ATGGGGCTCCCTCGTGACCTCTCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 113
 QY 21 CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
 Db 114 TGGCGGGCTCTCGAGCGGTGGCGGGCTCTTCAGGAGCGTGAAGTGAACCTTGAGGGG 173
 QY 41 GlyGlyAlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
 Db 174 GGAGGGCGGAGCAGGAGCGCGCCAGCGCTGGGAAAGTATTCTATGGCTGCCCTGGG 233
 QY 61 GlnGluProAlaLeuPheSerThrAspAsnAspAspPheThrValArgAsnGlyGluThr 80
 Db 234 CAAGAGCCAGCTCTGTATTAGCACTGATATATGATGACTTCACCTGCGGGAATGGCGAGA 293
 QY 81 ValGlnGluArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
 Db 294 GTCCAGAAAGAGGTCACTGAAGGAAAGGAATCCATGGAAGATCTTCCATCCAAAGCT 353
 QY 101 IleLeuArgArgHisLysArgAspTrpValAlaProIleSerValProGluAsnGly 120
 Db 354 ATCTTACGAGACACAAAGAGATTGGGTGGTGTCTCCAAATATCTGTCTCCCTGAAATGGC 413
 QY 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
 Db 414 AAGGTCTCTCTCCCGAGAGCTGATCAGCTCAAGTCTAATAAGATAGAGACCAAG 473
 QY 141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProGlyGlyValPheAlaVal 160
 Db 474 ATTTTCTACAGCATCACGGGGCGGGGCGAGACAGCCCTCTGAGGGTGTCTTCGTGTA 533
 QY 161 GluLysGluThrGlyTrpLeuLeuLeuLysProLeuAspArgGluLeuLeuLys 180
 Db 534 GAGAGGAGACAGGCTGGTGTGTGTAATAAGCCACTGGACCGGAGAGATTGCCAAG 593
 QY 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
 Db 594 TATGAGCTCTTTGGCCACGCTGTGTGAGAGATGTGCTCAGTGGAGGACCCCATGAAC 653
 QY 201 IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
 Db 654 ATCTCCATCATCGTACCGACAGCAATGCCAAGCCCAAGTTTACCAGGACACCTTC 713
 QY 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
 Db 714 CGAGGGAGTGTCTTAGAGGGAGTCTTACAGGTACTTCTGTGATCGAGGTGACCCACA 773
 QY 241 AspGluAspAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln 260
 Db 774 GATGAGATGATGCCATCTACACCTTACAATGGGTGGTGTCTTACTCCATCCATAGCCAA 833
 QY 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
 Db 834 GAAACCAAGGACCCACAGCCTTCAATTCACCAATCCCGAGACAGGACCATCAGC 893
 QY 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
 Db 894 GTCATCTCCAGTGGCTGGACCGGGAAAAGTCCCTCGTAGTACACACTGACCATCCAGGCC 953
 QY 301 ThrAspMetAspGlyAspGlySerThrThrThrAlaValAlaValGluLeuLeuAsp 320

Db 114 TCGCGCGCTCGAGCCGTGCGGCGGTCTTCAGGGAGGCTGAAGTGACCTTGAGGCG 173
Qy 41 GlyGlyAlaGluGlnProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
Db 174 GAGGCGCGGAGCAGGAGCCCGGCGGCGCTGGGAAAGTATTTCATGGGCTGCCCTGGG 233
Qy 61 GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr 80
Db 234 CAAGAGCCAGCTCTGTTTAGCACATGATATGATGACCTTCACTGTGCGGAATGGCGAGCA 293
Qy 81 ValGlnGluArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
Db 294 GTCCAGGAAGAAGTCACTAGGAAAGGAATCCATTAAGATCTTCCATCCAAAGT 353
Qy 101 IleLeuArgArgHisLysArgAspTrpValValAlaProIleSerValProGluAsnGly 120
Db 354 ATCTTACGACACACAGAGAGATGGGTGTCTTCAATATCTGTCCCTTGAAAAATGGC 413
Qy 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
Db 414 AAGGGTCCCTTCCCCAGAGACTGAATCAGCTCAAGTCTATAAAGATAGAGACCAAG 473
Qy 141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProProGluGlyValPheAlaVal 160
Db 474 ATTTTCTACAGCATCAGGGGCGGGGCGGACAGAGCCCTCGAGGGTGTCTTCGCTGTA 533
Qy 161 GluLysGluThrGlyTrpLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys 180
Db 534 GAGAGAGAGACAGCTGTTGTTGAATAAGCCACTGGACCGGGAGGAGATGGCAAG 593
Qy 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
Db 594 TATGAGCTCTTTGGCCACGCTGTGTACAGAGAAATGGTGCCTCAGTGGAGGACCCCATGAAC 653
Qy 201 IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
Db 654 ATCTCCATCATGTGACCGACCAAGATGACCAAGCCCAAGTTTACCCAGGACACCTTC 713
Qy 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
Db 714 CGAGGGAGTGTCTAGAGGGAGTCTTACAGGTAATCTGTGATGAGGTGACAGCCACA 773
Qy 241 AspGluAspAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln 260
Db 774 GATGAGGATGATGCATCTACACCTACATGGGGTGTGTCTTACTCCATCCATAGCCAA 833
Qy 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
Db 834 GAACCAAGAGACCCACACGACCTCATGTTCAAAATTCACCGAGCAGGACCATCAGC 893
Qy 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
Db 894 GTCATCTCCAGTGGCTGGACCGGAAAGTCCCTGAGTACACACTGACCATCCAGGCC 953
Qy 301 ThrAspMetAspGlyAspGlySerThrThrAlaValAlaValValGluIleLeuAsp 320
Db 954 ACAGACATGATGGGACCGCTCCACACACCGGAGTGGCAGTAGTGAGATCTTGTAT 1013
Qy 321 AlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
Db 1014 GCCAATGCAATGCTCCATGTTTGACCCCCAGAGTACAGGCGCCATGTGCTGAGAT 1073
Qy 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
Db 1074 GCAGTGGCGCATGAGGTGACAGGCTGACCGGTCACTGATCTGACGCGCCCACTCACCA 1133
Qy 361 AlaTrpArgAlaThrTyrLeuIleMetGlyValAspAspGlyAspHisPheThrIleThr 380
Db 1134 GCGTGGCGTCCACCTTACCTTATCATGGGCGGTGACAGCGGGACCATTTTACCATCAC 1193
Qy 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400

Db 1194 ACCACCCCTGAGAGCAACCCAGGCGCATCTTGACAAACGAGGAAGGTTTGGATTTTGAGGCC 1253
Qy 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db 1254 AAAAAACAGACACACCTGTAGCTTGAAGTGACCAACGAGGCCCTTTTGTGCTGAAGCTC 1313
Qy 421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
Db 1314 CCAACCTCCACAGCCACCATAGTGTCTCCAGCTGGAGGATGTGAATGAGGCACCTGTGT 1373
Qy 441 ValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
Db 1374 GTCCCACTCCAAAGTCTTGAAGTCCAGAGGGGATCCCACTGGGAGGCTGTGTGT 1433
Qy 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db 1434 GTCTACACTGCAGAGACCCCTCAAGGAGAAATCAAAAGATCAGCTACCGCATCCTGAGA 1493
Qy 481 AspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db 1494 GACCCAGCAGGGTGGCTAGCCATGGACCCAGACAGTGGGCAGGTCACTGTGGGCACC 1553
Qy 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db 1554 CTCGACCGTGAAGTGAAGTGTGTGAGGAACAACATCTATGAAGTCATGCTTTGGCC 1613
Qy 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuLeu 540
Db 1614 ATGGCAATGGAGAGCCCTCCCACTGGCAGCGGAAACCCCTTCTGCTAACACTGATTGAT 1673
Qy 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 1674 GTCAACGACCATGGCCAGCTCCCTGAGCCCTGTGTGAGATCACCATTGTGCACCAAGCCCT 1733
Qy 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
Db 1734 GTGCGCCAGCTGTGAACATCACGACAGGACCTGTCTCCCAACACCTCCCTCCCTTCCAG 1793
Qy 581 AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGlyAsp 600
Db 1794 GCGCAGCTCACAGATGATCTAGCATCTACTGGCGGAGAGGTCAACGAGGAAGGTGAC 1853
Qy 601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
Db 1854 ACAGTGGTCTTCTCCCTGAAGAAGTCTCTGAAGCAGGATACATATGACGTGCACCTTCT 1913
Qy 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
Db 1914 CTGTCTGACCATGGCAACAAAGAGCAGCTGACGGTGTATCAGGGCCACTGTGTGCGACTGC 1973
Qy 641 HisGlyHisValGluThrCysProGlyProTrpLysGlyGlyPheIleLeuProValLeu 660
Db 1974 CATGGCCATGTGAAACCTGCCCTGGACCTGGAAAGGAGGTTTTCATCTCCCTGTGCTG 2033
Qy 661 GlyAlaValLeuAlaLeuPheLeuLeuValLeuLeuLeuValArgLysLys 680
Db 2034 GGGGCTGTCTGGCTCTCTGTTCTCTCTGCTGTGCTGTCTTGTGCTGAGAAAGAA 2093
Qy 681 ArgLysIleLysGluProLeuLeuLeuProGluAspThrArgAspAsnValPheTyr 700
Db 2094 CGAGAGATCAGGAGCCCTCTACTCCAGAGATGACACCCCGTGACAGCTCTTCTTAC 2153
Qy 701 TyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
Db 2154 TATGCGAAGAGGGGGTGGCGAGAGAGACGAGGATATGACATCACCAGCTCCACCGA 2213
Qy 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
Db 2214 GGTCTGGAGGCGGAGGCTGCTTCCGCAATGACGTGGGCGGACCAACCATCATCCCG 2273
Qy 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
Db 2274 ACACCATGTACCGTCTTAGGCCAGCCAAACCCAGATGAATCGCAACTTTTATAATTGAG 2333

[illegible]

Db	1074	GCAGTGGCCATGAGGTGAGGCTGACGGTCACTGATCTGGACGCCCACTCACCA	1133
Qy	361	AlaTrpArgAlaThrTyLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr	380
Db	1134	CGGTGGCGTGCACCTACCTTATCATGCGCGTGCAGCGGGACCATTTTACCATCAC	1193
Qy	381	ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla	400
Db	1194	ACCCACCTGTGAGAGCAACAGGGCATCTGACCAACAGGAAGGGTTTGGATTTTGGGCC	1253
Qy	401	LysAsnGlnHisThrLeuTyValGluValThrAsnGluAlaProPheValLeuLysLeu	420
Db	1254	AAAAACACGACACACCTGTACGTTCAGTACCAACAGAGGCCCTTTTGTGTGAAGCTC	1313
Qy	421	ProThrSerThrAlaThrIleValHisValGluAspValAsnGluAlaProValPhe	440
Db	1314	CCAAACCTCCACAGCCACCATAGTGTCCACGTGGAGGATGTGAATGAGGCACCTGTGTT	1373
Qy	441	ValProProSerLysValValGluValGlnGluIleProThrGlyGluProValCys	460
Db	1374	GTCCACCTCCAAAGTGTGTGAGTCCAGAGGGCATCCCATCTGGGAGCCCTGTGTGT	1433
Qy	461	ValTyThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyArgIleLeuArg	480
Db	1434	GTCTACACTGCAGAGAACCTCTGACAGAGAAATCAAAAGATCAGCTACCGCATCTCGA	1493
Qy	481	AspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr	500
Db	1494	GACCCAGAGGTGGCTAGCCATGGACCCAGACAGTGGGAGGFCACAGCTGTGGGCACC	1553
Qy	501	LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyGluValMetValLeuAla	520
Db	1554	CTCGACCGTGAAGATGAGCAGTTTGTGAGGACCAACATCATGAGTCATGCTTGGCC	1613
Qy	521	MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAsp	540
Db	1614	ATGGACAAATGGAACCCCTCCACCATCGGCAGCGGAAACCCCTCTGTCTTAACACTGAT	1673
Qy	541	ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro	560
Db	1674	GTCAACCAACCATGGCCCGCCCTGAGCCCGCTGAGTACCATCATCTGCAACCAAGCCCT	1733
Qy	561	ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln	580
Db	1734	GTGGCCACGTGTGTGAACATCATCGGACAGGACCTGTCTCCACACCTCCCTTCCAG	1793
Qy	581	AlaGlnLeuThrAspAspSerAspIleTyTrpThrAlaGluValAsnGluGlyAsp	600
Db	1794	GCCAGCTCACAGATGACTCAGACATCTCTGACGCGCAGAGTCAACAGGAGAGTGAC	1853
Qy	601	ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyAspValHisLeuSer	620
Db	1854	ACAGTGTCTGTCTCCCTGGAAGAGTTCCTCAAGCAGGATACATATGACGTGCACCTTCT	1913
Qy	621	LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys	640
Db	1914	CTGTCTACCATGCGCAACAAGACAGCTACCGTGTATCAGGCCCATCTGTGTGCGACTGC	1973
Qy	641	HisGlyHisValGluThrCysProGlyProTrpLysGlyPheIleLeuProValLeu	660
Db	1974	CATGGCATGTGCAACCTGCTCCCTGACCTCGAAGAGAGTTTCATCTCTCTGTGCTG	2033
Qy	661	GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysLys	680
Db	2034	GGGGCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT	2093
Qy	681	ArgLysLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTy	700
Db	2094	CGAAGATCAAGAGGGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	2153
Qy	701	TyrGlyGluGlyGlyGlyGluGluAspGlnAspTyAspIleThrGlnLeuHisArg	720

Db	2154	TATGGCGAAGAGGGGGTGGCGAAGAGGACCGAGGACTATGACATCACCGAGTCCACCGA	2213
Qy	721	GlyLeuGluAlaArgProGluValValLeuAsnAspValAlaProThrIleIlePro	740
Db	2214	GGTCTGGAGGCCAGCGCGAGGTGGTTCTCCGCAATGAGCTGGCGACCAACCATCATCCG	2273
Qy	741	ThrProMetTyArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu	760
Db	2274	ACACCATGTACCGTCTAGGCGAGCAACCCAGATGAAATCGCAACTTTATAATTGAG	2333
Qy	761	AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyAspThrLeuLeuValPhe	780
Db	2334	AACCTGAAGCGGCTAACACAGACCCCGCCCTAGACACCTCTTGTGGTGTTC	2393
Qy	781	AspTyGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer	800
Db	2394	GACTATGAGGCGAGCGGCTCCGAGCGCGTCCCTGAGCTCCCTCACCTCTCCGCTCC	2453
Qy	801	AspGlnAsnGlnAspTyArgProGluValLeuAsnGluTrpGlySerArgPheLysLeuAla	820
Db	2454	GACCAAGACCAAGATTACGATTATCTGAACGAGTGGGCGAGCGCTTCAAGAAGCTGGCA	2513
Qy	821	AspMetTyGlyGlyGlyGluAspAsp	829
Db	2514	GACATGTACGTTGGCGGAGGAGGACGAC	2540
RESULT 5	BC041846	3686 bp	linear
LOCUS	BC041846		
DEFINITION	Homo sapiens cadherin 3, type 1, P-cadherin (placental), mRNA (CDNA clone MGC:43574 IMAGE:5269523), complete cds.		
ACCESSION	BC041846		
VERSION	BC041846.1	GI:27693123	
KEYWORDS	MGC.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	1 (bases 1 to 3686) Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo. Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.		
AUTHORS	Straussberg, R.L., Feingold, B.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Schetz, T.E., Brownstein, M.J., Usdin, T.B., Toshyuki, S., Carrinci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalish, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.		
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	99 (26), 16899-16903	(2002)
MEDLINE	22388257		
PUBMED	12477932		
REFERENCE	2 (bases 1 to 3686) Strausberg, R.		
AUTHORS	Direct Submission		
TITLE	Submitted (23-DEC-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
JOURNAL	NIH-MGC Project URL: http://mgc.nci.nih.gov		
REMARK	Contact: MGC help desk		
COMMENT	Email: cgapbs-remail.nih.gov		
	Tissue Procurement: Miklos Palkovits, M.D., Ph.D.		

QY 361 AlaTyrArgAlaThrTyrLeuIleMetGlyGlyAspGlyAspHisPheThrileThr 380
Db 1625 GGTGCGGTGCCACCTACCTATCATGCGCGGTGACGACGGGAGCAATTTACCATCACC 1684
QY 381 ThrHisProGluSerAsnGlnGlyIleThrThrArgLysGlyLeuAspPheGluAla 400
Db 1685 ACCACCTTCAGAGCAACAGGGGATCCCTGCAACACAGAGAGGGTTTGATTTTGAGGCC 1744
QY 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db 1745 AAAAACCCAGCACACCTGTACGTTGAAGTGACCAAGAGGCCCTTTTGTGCTGAGCTC 1804
QY 421 ProThrSerThrAlaThrIleValHisValGluAspValAsnGluAlaProValPhe 440
Db 1805 CCAACCTTCACAGCACCATATAGTGTCCAGCTGGAGGATGTGAATGAGGCCACCTGTGTT 1864
QY 441 ValProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
Db 1865 GTCCACCTCCAAAGTGTGTGAGTCCAGAGGGCATCCCACTGGGAGCCCTGTGTGT 1924
QY 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db 1925 GTCTACCTCAGAGAGACCTGACAAAGAGATCAAAAGATCAGCTACCGCATCTCGAGA 1984
QY 481 AspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db 1985 GACCCAGCAGGGTGTGCTAGCATGATGACCAAGAGAGTGGCGAGGTACAGCTGTGGGACC 2044
QY 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db 2045 CTCGACCTGAGGATGACAGTGTGTGAGGAAACAACATCATGAGTCAATGCTGTGGCC 2104
QY 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAsp 540
Db 2105 ATGGACAAATGAAGCCCTCCCACTGGCAGGAGAACCTTCTGTCAACTGATTCAT 2164
QY 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 2165 GTCAACGACCATGCGCCAGTCCCTGAGCCCGTCCAGATCACCATCTGCACCAAGCCCT 2224
QY 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
Db 2225 GTGGCGCAGGTGCTGAACATCAGCAGCAAGAGACCTGTCTCCCACTCCCTTCCAG 2284
QY 581 AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGlyAsp 600
Db 2285 GCCCAGCTCAGATGACTCAGACATCTACTGGACGGCAGAGGTCAACAGGAGGTGAC 2344
QY 601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
Db 2345 ACAGTGTCTTGTCCCTGAAGAACTTCTGAAGCAGGATACATATGACGTGACCTTTCT 2404
QY 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
Db 2405 CTGTCTGACCATGTCACAAAGAGCAGCTACCGTGTATCAGGGCCACTGTGTCCGACTGC 2464
QY 641 HisGlyHisValGluThrCysProGlyProTyrLysGlyGlyPheIleLeuProValLeu 660
Db 2465 CATGGCCATGTGAAACCTGCTCCCTGACCTCGAAGAGGTTTCATCCCTCTGTGCTG 2524
QY 661 GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysLys 680
Db 2525 GGGGCTGTCTGCTGTCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2584
QY 681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
Db 2585 CGAAGATCAAGAGGCCCT 2644
QY 701 TyrGlyGluGlyGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
Db 2645 TATGGGCAAGAGGGGGTGGCAAGAGGACCAAGGACTATGACATCACCAGCTCCACCA 2704

721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
Db 2705 GGTCTGAGGCCAGCGCGGAGGTGTTCTCGCAATCAGCTGGCACCACCATCATCCCG 2764
QY 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleGlu 760
Db 2765 ACACCCATGTACCGTCTAGCGCAGCCACCCAGATGAAATCGCAACTTTATAATTGAG 2824
QY 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
Db 2825 AACCTGAAGCGCGCTAACACAGACCCACAGCCCGCCCTTACCTCTCCGCTCC 2884
QY 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
Db 2885 GACTATGAGGCGAGCGCTCCGACGCGCGTCCCTGAGCTCCCTACCTCTCCGCTCC 2944
QY 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTyrGlySerArgPheLysLysLeuAla 820
Db 2945 GACCAAGACCAAGATTACGATTATCTGAACAGTGGGCGAGCGCTTCAAGAGCTGGCA 3004
QY 821 AspMetTyrGlyGlyGlyGluAspAsp 829
Db 3005 GACATGTACGCTGCGGGGAGGACGAC 3031
RESULT 6
AX382118
LOCUS
DEFINITION
Sequence 17 from Patent WO0200939.
ACCESSION
AX382118
VERSION
AX382118.1 GI:19576927
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1
AUTHORS
Macina, R.A. and Pillai, R.
TITLE
Method of diagnosing, monitoring, staging, imaging and treating
colon cancer
JOURNAL
Patent: WO 0200939-A 17 03-JAN-2002;
Diadexus, Inc. (US)
FEATURES
Location/Qualifiers
1..3219
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 3219
Score: 4348.00 Matches: 828
Percent Similarity: 99.76% Conservative: 0
Best Local Similarity: 99.76% Mismatches: 1
Query Match: 99.52% Indels: 1
DB: 6 Gaps: 0
US-09-916-849A-1 (1-829) x AX382118 (1-3219)
QY 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuLeuValCysTrpLeuGln 20
Db 90 ATGGGGCTCCCTCGTGGACCTCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 149
QY 21 CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
Db 150 TCCGGCGCTCCGACCGCTCGGCGGCTCTTCAGGAGGCTGAAGTACCTTGGAGCG 209
QY 41 GlyGlyAlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
Db 210 GGAGCGCGGAGCAGGAGCCCGCAGCGCTGGGAAAGTATTTCATGGCTGCCCTGGG 269
QY 61 GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr 80
Db 270 CAAGAGCCAGCTCTGTTTACGACTGATGATGACTTCACTGCGGAATGGGAGACA 329

Qy 81 ValGlnGluArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
Db 330 GTCCAGGAAGAGGTCACCTGAAGGAAGGAATCCATTGAAGATCTTCCCATCCAAACGT 389
Qy 101 IleLeuArgArgHisLysArgAspTropValValAlaIleProIleSerValProGluLysGly 120
Db 390 ATCTTACGAAGACACAGAGAGATTGGGTGGTGTCTCAATATCTGTCCCTGAAATGGC 449
Qy 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
Db 450 AAGGTCCCTTCCCCAGAGACTGAATCAGCTCAAGTCTTAATAAGATAGAGACACCAAG 509
Qy 141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProGluGlyValPheAlaVal 160
Db 510 ATTTTCTACAGCATCACGGGGCGGGGCGAGACGCCCCCTCGAGGGTGTCTTCGCTGTA 569
Qy 161 GluLysGluThrGlyTyrLeuLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys 180
Db 570 GAGAAGGAGACAGGCTGGTGTGTGTAATAAGCCACTGGACCGGGAGAGATTGCCAAG 629
Qy 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
Db 630 TATGAGCTCTTTGGCCACCGCTGTGCAGAGAAATGGTGCCTCAGTGGAGACCCCATGAAC 689
Qy 201 IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
Db 690 ATCTCCATCATCTGACCCAGCAGATGACCAAGCCAAAGTTTATCCAGGACACCTTC 749
Qy 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
Db 750 CGAGGGAGTGTCTAGAGGAGTCTTACCAGGTACTTCTGTGATGACAGGTGACAGCCACA 809
Qy 241 AspGluAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln 260
Db 810 GATGAGGATGATGCCATCTACACCTACAAATGAGGGTGGTGTCTTACTCCATCATAGCCAA 869
Qy 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
Db 870 GAACCAAGAGACCCACACGACTCATGTTCACAAATCCACGGAGACAGGACCATCAGC 929
Qy 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
Db 930 GTCATCTCCAGTGGCTGACCCGGGAAAAGTCCCTGAGTACACACTGACCATCCAGGCC 989
Qy 301 ThrAspMetAspGlyAspGlySerThrThrThrAlaValAlaValValGluIleLeuAsp 320
Db 990 ACAGACATGGATGGGACCGGCTCCACCCACCGGAGTGGCAGTAGTGGAGATCCTTGAT 1049
Qy 321 AlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
Db 1050 GCCAATGACATGCTCCCATGTTTACCCCCAGAGAGTACGAGGCCCATGTGCTGAGAA 1109
Qy 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
Db 1110 GCAGTGGGCGATGAGGTGAGAGGCTGACGGTCACTGATCTGGACGCCCCCACTCACCA 1169
Qy 361 AlaTyrArgAlaThrThrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
Db 1170 GCGTGGCGTGCACCTTACCTTATATGCGGTGACGACGCGGGACCATTTTATCCATCAC 1229
Qy 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
Db 1230 ACCACCCCTGAGAGAACAGGGCATCTCTGACACACGAGAGGGTGTGGATTTTGAGGCC 1289
Qy 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db 1290 AAAAACCCAGCACACCTGTACCTTGAAGTGAACCAACGAGGGCCCTTTTGTGCTGAAGCTC 1349
Qy 421 ProThrSerThrAlaThrIleValHisValGluAspValAsnGluAlaProValPhe 440
Db 1350 CCAACCTCCACAGCCACCATAGTGTCTCCAGCTGGAGGATGTAATGAGGACCTGTGTGT 1409

Qy 441 ValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
Db 1410 GTCCCAACCTCTCAAAAGTGTGTGAGGTCCAGGAGGCATCCCACTGGGGAGCCTGTGTGT 1469
Qy 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrAsnIleLeuArg 480
Db 1470 GTCTACATGTCGAGAAGACCTTGACAGAGAGATCAAAAGATCAGCTACCCGATCCTGAGA 1529
Qy 481 AspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db 1530 GACCAGCAGAGGTGTGCTAGCCATGGACACACAGTGGCAGGTACACAGTGTGGGCACC 1589
Qy 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db 1590 CTCGACCGTGGAGATGACAGTTTGTGAGGAAACAATCTATGAGATCATGGTCTTGGCC 1649
Qy 521 MetAspAsnGlySerProProThrThrGlyThrLeuLeuLeuThrLeuLeuLeuAsp 540
Db 1650 ATGGCAATGGGAAGCCCTCCACCACTGGCAGCGAAACCTTCTCTGAACACTGATTGAT 1709
Qy 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 1710 GTCAACGACCATGGCCCGCTCCCTGAGCCCGCTCAGATCACCATCTGCAACCAAGCCCT 1769
Qy 561 ValArgHisVal-LeuAsnIleThrAspLysAspLysSerProHisThrSerProPheGln 580
Db 1770 GTGCGCCAGGTGGTGAACATCAGCGACAGGACCTGTCTCCCAACACTCCCTTCCA 1829
Qy 580 AlaGlnLeuThrAspAspSerAspIleTyrThrAlaGluValAsnGluGluLys 600
Db 1830 GGCCAGCTCACAGATGACTCAGACATCTACTGACCGGACAGGTCAACGAGGAAGTGA 1889
Qy 600 pThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSe 620
Db 1890 CACAGTGGTGTGCTGCTGAAGAAGTTCCTGAAGCAGGATACATATGACGTGACCTTTC 1949
Qy 620 rLeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
Db 1950 TCTGTCTGACCATGGCAACAAAGAGAGCTGACGGTGATCAGGGCCACTGTGTGCGACTG 2009
Qy 640 sHisGlyHisValGluThrCysProGlyProTyrPlyGlyGlyPheIleLeuProValle 660
Db 2010 CCATGGCCATGTGCAAAACCTGCCCTGGACCTGGAAGGAGGTTTCATCTCCCTGTGCT 2069
Qy 660 sGlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysLys 680
Db 2070 GGGGCTGTCTGCTGTCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2129
Qy 680 sArgLysIleLysGluProLeuLeuProGluAspAspThrArgAspAsnValPheTy 700
Db 2130 GCGGAAGATCAAGAGACCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2189
Qy 700 rTyrGlyGluGluGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisAr 720
Db 2190 CTATGGCGAAGAGGGGTGGCGAAGAGACAGGACTATGACATCACCGAGTCCACCG 2249
Qy 720 gGlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePr 740
Db 2250 AGGTCTGAGGCCAGGCGGAGGTGTTCTCCGCAATGACGTGGCAGCAACCATCATCC 2309
Qy 740 rThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGl 760
Db 2310 GACACCCATGATCGTCTTAGCCGAGCAACCCAGATGAATCGGCAACTTATTAATTGA 2369
Qy 760 uAsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPh 780
Db 2370 GAACCTGAAGCGCGCTAACACAGACCCCAAGCCCGCCCTAGCAGCACCTCTTGTGT 2429
Qy 780 sAspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSe 800
Db 2430 CGACTATGAGGGCAGCGGCTCCGACGCGGCTGCTGAGCTCTCTCACCTCTCCGCCCTC 2489
Qy 800 rAspGlnAspGlnAspTyrAspTyrLeuAsnGluTyrpGlySerArgPheLysLysLeuAl 820

Db 2490 CGACCAAGACCAAGATTACGATTATCTGAACAGCTGGCGGACCGCTTCTAGAACGCTGGC 2549

Qy 820 aAspMetTyrGlyGlyGlyGluAspAsp 829

Db 2550 AGACATGTACGGTGGCGGGAGGACGAC 2577

RESULT 7

MMCADHP

LOCUS

DEFINITION Mouse mRNA for P-cadherin.

ACCESSION X06340

VERSION X06340.1 GI:50267

KEYWORDS cadherin; cell adhesion molecule; glycoprotein; transmembrane protein.

SOURCE

ORGANISM Mus musculus (house mouse)

REFERENCE

AUTHORS Nose.A., Nagafuchi.A. and Takeichi.M.

TITLE Isolation of placental cadherin cDNA: identification of a novel gene family of cell-cell adhesion molecules

JOURNAL EMO J. 6 (12), 3655-3661 (1987)

MEDLINE 88111554

PubMed 3428270

FEATURES

source

1..3187

/Location/Qualifiers

/organism="Mus musculus"

/mol_type="mRNA"

/db_xref="taxon:10090"

/clone="p3"

/cell_line="PSA5-E"

46..2514

/note="unnamed protein product; pre P-cadherin (AA -27 to 793)"

/codon_start=1

/protein_id="CAA29646.1"

/db_xref="GI:50268"

/db_xref="GOA:P10287"

/db_xref="SWISS-PROT:P10287"

/translation="MELLSGPHAFLLLLQVCLRWSEPYRAGFIGAGVTLVEVG
TDLPEQVGLKVALAGCGKHHADNGDIIMLTGTGGKADMSHSPETRLIRKREW
MPPIFVPEKNGKFFQRLNQLKNDGRGKIFYSITPGADSPPEGVFTIEKSGWLL
LHMLPDLREKIVKYELCHAVSNGASVEEPMNISIIVTDNDKPKFTODTPRGSVLE
GMPGTSQVQVATDDEDDAVNTGVVAYSISHOEPKPHDLMTIHKSTGTISVSS
GLDRKXVPEYLTQVATDMDGEGSTTTAAVQLILDANDANPEFEPQKRYEAVPNEVY
GHEVQLTVTDLDVNPWAPWATYHIVGGDGDGHFTIITHPSTNQVLTKKGLDFEA
QOQHTLVETNEAPFAVLPTATATVHVGDVNEAPVFPSPKRYEAEQGISIGEL
VCYITAQDPEKQKISYTSRDPANWLDPDSQITAAGLDREDEQFVKNNYVEV
MWLADSNPPTGTGTLTLLTDINDHGPIDPEPQIITCNQSPVQVLNITDRKLS
NNSPFAQIITHSDIYVMAEVSKEGTVALSKFLKQDTYDLHLSDHGNREQLTM
IRATVCDCHQVFNDCPRWKGFILPQLCAVLAITLLALLHLVRKKRYKPEKPLL
PEDTDRNIFYYEGEGGEDDYDITQLHGLEARPEVVLNDVVPFIPTPYRPR
PANPDIGINLIENLKPAKTDTPAPYDSLWFDYEGSGSDAASLSLITTSASDQDQD
YNYLWEGSRFKFLADNYGGGDD"

46..126

/note="signal peptide (AA -27 to -1)"

127..2511

/product="P-cadherin (AA 1 - 795)"

619..621

/note="pot. N-linked glycosylation site"

1690..1692

/note="pot. N-linked glycosylation site"

1717..1719

/note="pot. N-linked glycosylation site"

1744..1746

/note="pot. N-linked glycosylation site"

3164..3170

/note="polyA signal"

ORIGIN

alignment Scores:

Qy	319	LeuAspAlaAsnAspAsnAlaProMetPheAspProGlnLysTyrgluAlaHisValPro	338
Db	976	CTTGATGCCAACGATAACCGCTCCGGAGTGTGGCCGACAGAGTATGAGGCTTGGGTGGCT	1035
Qy	339	GluAsnAlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsn	358
Db	1036	GAGACCGAAGTGGGCCATAGGTCACAGGCTCAGAGTGCATCTCGATGTCCTCCCAAC	1095
Qy	359	SerProAlaTtpArgAlaThrTyLeuIleMetGlyGlyAspAspGlyAspHisPheThr	378
Db	1096	TGGCCACGCGTGGCGTCCACCATTACCATCGTGGAGGTGATGATGGGACCATTTCCACC	1155
Qy	379	IleThrThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPhe	398
Db	1156	ATCACCACTCACCCAGAGACCCACRAGGGGTCTGACCAACCAAGAGGGTTGGATTTT	1215
Qy	399	GluAlaLysAsnGlnHisThrLeuTyValGluValThrAsnGluAlaProPheValLeu	418
Db	1216	GAGGCTTCAGGACCAACACACCCTGTATGTAGAGTGCACCAAGGAGCTCCCTTTGCAGT	1275
Qy	419	LysLeuProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaPro	438
Db	1276	AACTCCCGACAGCCACTGCCACGCTGGTCCATGTGAAGATGTCAACGAGCCCT	1335
Qy	439	ValPheValProSerLysValValGluValGlnGluGlyIleProThrGlyValPro	458
Db	1336	GTGTTGTGTCCACTTCCACAGTCATTGAGGCCCGAGGAAGGCATCTCTATTGGGAACTG	1395
Qy	459	ValCysValTyThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyArgIle	478
Db	1396	GTCTGCATCTATCCGCACAGGCCACAGACCAAGAGGACCAAGATCAGCTACACCATC	1455
Qy	479	LeuArgAspProAlaGlyTtpLeuAlaMetAspProAspSerGlyGlnValThrAlaVal	498
Db	1456	TCGAGAGATCCAGCCAACTGGCTTGTGTGGACCCACAGACAGTGGTCAGATACTCGCGCA	1515
Qy	499	GlyThrLeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyxGluValMetVal	518
Db	1516	GGCATCTTTGGATCGTAGGACGAGCAGTTTGTGAAAAACAATGTCTACGAAGTCATGGTT	1575
Qy	519	LeuAlaMetAspAsnGlySerProProThrGlyThrGlyThrLeuLeuLeuThrLeu	538
Db	1576	TTGGCCACAGACAGTGGAAACCTCCCAACCACCGCAGCTGGACCCCTCTGCTTACACTT	1635
Qy	539	IleAspValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGln	558
Db	1636	ACTGCATCAACACCATGGCCCGATCCCTGAAACCCAGGCAGATCATCATCTGTAAACAA	1695
Qy	559	SerProValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerPro	578
Db	1696	AGCCCTGTGCCTCAAGTCTGAACATCAGTCACAGAGACCTGTGCCCACTCTCCCTCCCT	1755
Qy	579	PheGlnAlaGlnLeuThrAspAspSerAspIleTyTtpThrAlaGluValAsnGluGlu	598
Db	1756	TTTCCAGGCCACAGCTAAACACATGATCAGATPATCTACTGGATGCCAGAAAGTCAGGAGAAA	1815
Qy	599	GlyAspThrValValLeuSerLeuLysPheLeuLysGlnAspThrTyxAspValHis	618
Db	1816	GGAGACACCGTGGCTTGTCCCTGAGAGAGTTCTCTGAAACAGACAGCTATGACTTCAT	1875
Qy	619	LeuSerLeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCys	638
Db	1876	CTTTCTTTGTCTGCCATGGCAACAGGGAACAGCTAACCATGATCAGGCGCACCTGTGTGT	1935
Qy	639	AspCysHisGlyHisVal---GluThrCysProGlyProTtpLysGlyGlyPheIleLeu	657
Db	1936	GACTGCCATGCCAAGTGTTCATGACTGCCCCACAGACCCCTGGAAGGGTGGTTTCATCTCTC	1995
Qy	658	ProValLeuGlyAlaValLeuAlaLeuLeuPheLeuLeuLeuValLeuLeuLeuVal	677
Db	1996	CCCATCTGGGTGTGTCTCGACATGCTGACCTTCTACTGGCACTCCTCCTCTGTGGTG	2055
Qy	678	ArgLysLysArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsn	697

Db	2056	ACGAGAAGAAAGGAAGTCAAGAGAGCCCTTCGTCCCGAAGATGACACGGAGACAAT	2115
QY	698	ValPheTYrTYrGlyGIuGIuGlyGIyGIuGIuAspGinAspTYrAspIleThrGln	717
Db	2116	GTCCTCATTATGAGAGAGAGGGTGTTGGTGAAGAGGACCAGGACTATGACATCACCCAA	2175
QY	718	IeuHisArgGlyLeuGIuAlaArgProGIuValValLeuArgAsnAspValAlaProThr	737
Db	2176	CCTCACCGGGGACTGGAGGCCAGCGCTGAGGTGGTTCCGAAAACGATAGTAGCCCAACC	2235
QY	738	IleIleProThrProMetTYrArgProArgProAlaAsnProAspGIuIleGlyAsnPhe	757
Db	2236	TTCATCCCCCCCATTACCGACCCCGCGCCCAACCAGATGAATCGGGAATTC	2295
QY	758	IleIleGIuAsnIeuIysAlaAlaAsnThrAspProThrAlaProProTYrAspThrLeu	777
Db	2296	ATCATCGAGAACCTGAAGCGCTGCACACTGACCTACTGCCCGCCTCAGACTCCCTG	2355
QY	778	IeuValPheAspTYrGIuGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSer	797
Db	2356	ATGGTTTTTGTACCTACGAGGCGCGGCTCTGATGCGGCTCCCTGAGTCCCTCACCA	2415
QY	798	SerAlaSerAspGlnAspGlnAspTYrAspTYrIeuAsnGIuTrpGIySerArgPheIys	817
Db	2416	TCCGCTCCGACACGAGTACGAGTACCACTACCTTAACGAGTGGGAGTGCATTC	2475
QY	818	LysLeuAlaAspMetTYrGIyGIyGIuAspAsp	829
Db	2476	AAATGCGGACATGTATGTCGCGTGAGGATGAC	2511
RESULT	8		
BC052189			
LOCUS	Mus musculus cadherin 3, mRNA (cdna clone IMAGE:6394535), partial cds.	2993 bp	mRNA linear ROD 16-SEP-2003
DEFINITION	Mus musculus		
ACCESSION	BC052189		
VERSION	BC052189.1	GI:30704940	
KEYWORDS			
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eutelestomii; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 2993) Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,I., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.P., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusik,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Uedin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raja,S.S., Joquellano,N.A., Peters,G.J., Abramson,R.D., Mullaly,S.J., Bosak,S.A., McSwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmuck,U., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smallos,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Warra,M.A. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)		
JOURNAL	23388257		
MEDLINE	PUBMED		
REFERENCE	2 (bases 1 to 2993)		
AUTHORS	Strausberg,R.		
TITLE	Direct Submission		
JOURNAL	Submitted (01-MAY-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,		

REMARK
COMMENT

FEATURES SOURCES

gene

CD5

```

KKLADMVGGGEDD"
105. .389
/gene="Cdh3"
/note="cadherin; Region: Cadherin domain"
/db xref="CDD:pfam00028"

```

```

/db_xref="CDD:p1am00028"
480.749
/gene="Cdh3"

```

/ gene- Cans
/note="CA; Region: Cadherin repeats. Cadherins are glycoproteins involved in Ca²⁺-mediated cell-cell adhesion. Cadherin domains occur as repeats in the

```
extracellular regions which are thought to mediate
cell-cell contact when bound to calcium"
/db xref="CDD:smart00112"
```

```

CDD:584200.12
768..1064
/gene="Cdh3"
/note="cadherin; Region: Cadherin domain"
/db_xref="CDD:pfam00028"
1158..1409
/gene="Cdh3"

```

```

1158. 1409. /Gene="Cdh3"
/Note="CA; Region: Cadherin repeats. Cadherins are
glycoproteins involved in Ca2+-mediated cell-cell
adhesion. Cadherin domains occur as repeats in the
extracellular regions which are thought to mediate
cell-cell contact when bound to calcium"
/db xref="CDD:smart00112"
1432. 1695

```

```

/ww_xref="CDD:Smart00112"
1422. .1685
/gene="Cdh3"
/not="cadherin; Region: Cadherin domain"
/db_xref="CDD:Pfam0028"
1845. .2252

```

1845. xref="CDD:pfam01049"
/gene="Cdh3"
/note="Cadherin_C_term; Region: Cadherin cytoplasmic region. Cadherins are vital in cell-cell adhesion during tissue differentiation. Cadherins are linked to the cytoskeleton by catenins. Catenins bind to the cytoplasmic tail of the cadherin. Cadherins cluster to form foci of homophilic binding units. A key determinant to the strength of the binding that it is mediated by cadherins is the juxta-membrane region of the cadherin. This region induces clustering and also binds to the protein p120ctn"
/db xref="CDD:pfam01049"

Alignment Scores:

Pred. No.:	3,77e-252	Length:	2993
Score:	343.50	Matches:	639
Percent Similarity:	94.01%	Conservative:	51
Best Local Similarity:	87.06%	Mismatches:	43
Query Match:	78.59%	Indels:	1
DB:	10	Gaps:	1

US-09-916-849A-1 (1-829) x BC052189 (1-2993)

Qy 97 ProSerLysArgIleLeuArgArgHisLysArgAspTrpValValAlaProIleSerVal 116

[illegible]

Db 60 CCACCCACCCGCATCTTAAGGAGACGAAAGAGAGAGTGGTGATGCCACCAATATTCGTC 119

QY
QY 117 ProGluAsnGlyLysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAsp 136

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

Db 120 CCGAGAA TGGCAAGGTCCCTTCAGAGGCTGAATCAGCTCAAATCTAATAAGGAC 179

137 ArgAsnThrIvsIlePheTyrSerIleThrGlvProGlvAlaAsnSerProProGlvGlv 156

QY 13 / A1GASPnrlYslllefnrYlserlllethrglypfroglYAlaaspserproglugly 156

Db 180 AGAGGCACCAAGATTTTCTACAGCATCACAGGGCCTGGCGCAGACAGTCCCCCGAAGGA 239

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

157 ValPheAlaValGluLysGluThrGlyTrpLeuLeuAsnLysProLeuAspArgGlu 176

D_b 240 GTCTTCACCATAGAGAAGGAGTCGGGCTGGCTGTTGCATATGCCACTGGACAGGGAG 299

177 GluIleAlaLysTyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGlu 196

db
300 AAGATTGTCAAGTACGAGCTTTATGGCCACCGCTGTATTCTGCACCATATCGCTCTCTACAC
::: ::::

500 AAGATGGTCAAGTACGAGCTTTATGGCCACGCTGTATCTGAGAAATGGTGGCTCTGTAGAG 359

QY 197 AspProMetAsnIleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThr 216

4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 5

Db 360 GAGCCCATGAACATATCCATCATTTGTGACAGACCAGAATGACAACAAGCCCCAGTTCACT 419

Qy 217 GlnAspThrPheArqGlySerValLeuGluGlyValLeuproGlyThrSerValMetGln 236

0
1
2
3
4
5
6
7
8
9
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y
Z

D**b** 420 CAAGACACCTTCAGAGGAGTGTCTGGAGGGAGTAATGCCCTGGCACTTCTGTGATGCAG 479

[illegible]

QY 23 / valinAlaThrAspGluAspAspAlaIleTyrIhrTyrASNGlyValValAlaTyrSer 256

[illegible]

Db 480 GTGACAGCCACAGATGAGGACGATGCTGTCAACACTTACAACTGGGTGGTCTTACTCC 539
 QY 257 ILeHisSerGlnGluProLysAspProHisAspLeuMetPheThrIleHisArgSerThr 276
 Db 540 ATCCATAGCCAGAGCGGAGGAGCCACACGACCTCATGTTCCATCCATCAATAAGACAGC 599
 QY 277 GlyThrIleSerValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeu 296
 Db 600 GGAACCATTAAGCGTCATATCCAGTGGCTGGACCGAGAGAAAGTCCCTGAGTACAGACTG 659
 QY 297 ThrIleGlnAlaThrAspMetAspGlyAspGlySerThrThrAlaValAlaVal 316
 Db 660 ACCGTCAGGCCACAGACATGAGTGGAGAGGCTCTTACCACCGCAGGAGCGCGTTGGT 719
 QY 317 GluIleLeuAspAlaAsnAspAlaProMetPheAspProGlnLysTyrGluAlaHis 336
 Db 720 CAAATCCTTGATGCCAACGATACGCTCCGGAGTTTGAGCCGACAGATGATGAGGCTTGG 779
 QY 337 ValProGluAsnAlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAla 356
 Db 780 GTCCCTGAGAACGAAGTGGCCATGAGGTACAGAGCTGACAGTCACTGATCTCGATGC 839
 QY 357 ProAsnSerProAlaTyrArgAlaThrTyrLeuIleMetGlyIleAspAspGlyAspHis 376
 Db 840 CCCAACTCGCCAGCGTGGCGTCCACTACCATCTGTTGGAGGTGATGATGGGACCAT 899
 QY 377 PheThrIleThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeu 396
 Db 900 TTCAACATCACCACTCACCAGAGACCAACCAAGCGCTCTGACAAACCAAGAGGGTTG 959
 QY 397 AspPheGluAlaLysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPhe 416
 Db 960 GATTTTGAGGCTCAGGACCAACACACCCCTGTATGAGAAGTACCAACAGAGGCTCCCTTT 1019
 QY 417 ValLeuLysLeuProThrSerThrAlaThrIleValValHisValGluAspValAsnGlu 436
 Db 1020 GCGATGAAGTCCCGACAGCCACTGCCCGCGTGGTCCATGTAAGATGTCAACGAA 1079
 QY 437 AlaProValPheValProSerLysValValGluValGlnGluGlyIleProThrGly 456
 Db 1080 GCCCTGTGTTGTTCCACTTCCAAAGTCACTGAGGCCCCAGGAGGCACTCTATTGGG 1139
 QY 457 GluProValCysValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyr 476
 Db 1140 GAATCGTGTGATCATCCACAGACCCAGACACAGGAGGACCAAGATCAGATCAGTAC 1199
 QY 477 ArgIleLeuArgAspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThr 496
 Db 1200 ACCATCTCGAGAGATCCAGCCAACTGGCTTGTGGAGCCACAGACAGTGGTCACTAAT 1259
 QY 497 AlaValGlyThrLeuAspArgLysGluAspGlnPheValArgAsnIleTyrGluVal 516
 Db 1260 GCGCAGGAGCATCTTGATCGTGAAGCAGGAGGAGTGTGAAAAACAATGTCTACGAAGTC 1319
 QY 517 MetValLeuAlaMetAspAsnGlySerProThrThrGlyThrGlyThrLeuLeuLeu 536
 Db 1320 ATGGTTTTGGCCACAGACAGATGAAACCTCCACCCAGCGCACTGGGACCTCCCTGCT 1379
 QY 537 ThrLeuIleAspValAsnAspHisGlyProValProGluProArgGlnIleThrIleCys 556
 Db 1380 ACATTTACTGACATCAACAGCAGCATGGCCGATGCCCTGAACCCAGGACATCATCTGT 1439
 QY 557 AsnGlnSerProValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThr 576
 Db 1440 AACCAAGCCCTGTGCTCAAGTGTGAACATCACTGACAGGACCTGTCCCCCACTCC 1499
 QY 577 SerProPheGlnAlaGlnLeuThrAspSerAspIleTyrThrAlaGluValAsn 596
 Db 1500 TCCCTTTTCCAGCCCGAGTAAACATCACTCAGATATCTACTGATGGGAGAGTCAAGC 1559
 QY 597 GluGluGlyAspThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAsp 616
 Db 1560 GAGAAAGGAGACACCGTGGCTGTGTCCTGAAGAAGTTCCTGAAACAGACACGATATGAC 1619

QY 617 ValHisLeuSerLeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThr 636
 Db 1620 TTGCATCTTCTTGTCTGACCACTGGCAACAGGAACAGCTAACCATGCTAGGCGCACT 1679
 QY 637 ValCysAspCysHisGlyHisVal---GluThrCysProGlyProTyrPlysGlyGlyPhe 655
 Db 1680 GTGTGTGACTGCTCATGGCCAAAGTGTCAATGACTGCCAGACCTCGAAGGGTGGTTTC 1739
 QY 656 IleLeuProValLeuGlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeu 675
 Db 1740 ATCTTCCCTCCTCTGGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1799
 QY 676 LeuValArgLysLysArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArg 695
 Db 1800 TTGGTGAG 1859
 QY 696 AspAsnValPheTyrTyrGlyGluGluGluGlyGlyGlyGluGluAspGlnAspTyrAsp 715
 Db 1860 GACATGCTCTTATATGAG 1919
 QY 716 ThrGlnLeuHisArgGlyLeuGluAlaArgProGluValValLeuArgAsnAspValAla 735
 Db 1920 ACCCAACTCCACCGGGAGTGGAGCCAGGCTGAGGTGGTCTCCGAAACGATGTAGTG 1979
 QY 736 ProThrIleLeuProThrProMetTyrArgProArgProAlaAsnProAspGluIleGly 755
 Db 1980 CCAACCTTCATCCCAACCCCATGTGTACGAGCCCGCCGCGCCCAACCATGATGAATCGG 2039
 QY 755 AsnPheIleIleGluAsnLeuLysAlaAlaAsnThrAspProThrAlaProTyrAsp 775
 Db 2040 AACTTTCATCATCGAGAACTGAAGCTGCCAACTGACCTACTGCGCCGCTACGAC 2099
 QY 776 ThrLeuLeuValPheAspTyrGluGlySerGlySerAspAlaAlaSerLeuSerLeu 795
 Db 2100 TCCCTGCTGGT 2159
 QY 796 ThrSerAlaSerAspGlnAspGlnAspTyrAspTyrLeuAsnGluTyrGlySerArg 815
 Db 2160 ACCACTCGCTCGACAGGATCAGACTACAACTTAACTTAACTTAACTTAACTTAACTTAA 2219
 QY 816 PheLysLysLeuAlaAspMetTyrGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 829
 Db 2220 TTCAAGAACTGGCGGACATGTATGGTGGCGGTGAGGATGAC 2261

RESULT 9
 GGCADHB
 LOCUS 2436 bp mRNA linear VRT 20-JUL-1995
 DEFINITION Chicken mRNA for B-cadherin.
 ACCESSION X58518
 VERSION X58518.1 GI.63113
 KEYWORDS B-cadherin; cadherin; transmembrane protein.
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
 REFERENCE 1 (bases 1 to 2436)
 AUTHORS Napolitano, E.W., Venstrom, K., Wheeler, E.F. and Reichardt, L.F.
 TITLE Molecular cloning and characterization of B-cadherin, a novel chick cadherin
 JOURNAL J. Cell Biol. 113 (4), 893-905 (1991)
 MEDLINE 91225083
 PUBMED 2026653
 REFERENCE 2 (bases 1 to 2436)
 AUTHORS Napolitano, E.W.
 TITLE Direct Submission
 JOURNAL Submitted (01-MAR-1991) E.W. Napolitano, Howard Hughes Medical Institute, University of California, 3rd and Parnassus Sts, San Francisco CA 94143-0724, USA
 FEATURES
 source location/Qualifiers
 1. 2436 /organism="Gallus gallus"

```

/mol_type="mRNA"
/db_xref="taxon:9031"
/clone="CAD-51"
/tissue_type="brain"
/clone_lib="lambda gt11 E13 chick brain cdna"
/dev_stages="embryonic day 13"
1..2436
/note="ORF1"
B-cadherin
/evidence=experimental
<1..2181
/note="ORF1"
/codon_start=1
/evidence=experimental
/product="B-cadherin"
/protein_id="CAA1408.1"
/db_xref="GI:63114"
/db_xref="GOA:P33145"
/db_xref="SWISS-PROT:P33145"
/translating="DWIPPIKVPENERGPPKLNVOIKSNRDEAKIFYISITQOGAD
APPGITIEKETGMKVTOPLDREHINKVLYSHVSENGKPEVERBMEIIVTVTON
DNKQFTQEVPRGSPVPGALPGTSMVRNATDADDDVETVNGVYAVLSOBEREPH
HMPYVNRATGTVSAGLDRVRVETLTQWADLDGQGLTTTALAVISITVDNNA
PEFQKTYEARAVPENEAEVARLATDDEPHTPAWRVYSIVRGNEGGAFTITDP
ASNEGLRTAKGLDYEARQFVHVAVNEAPFAIKLPTATATVMSVEDNEAPFD
PPLRAQVEDVPLQGLASVTAQDPDRAQQRIKYVMGSDPAGWLAVHPENGLIAR
BQLDRESFTKNSYVAVLAVDDGLPPTGTGTLTLLDNDHGPPEPRIVICN
RSPYVOLVITTDRLPNTGTPRAELSHGSDSWAVEVNGGDTVALWLEPELQNLV
SVLRLPDRGKQDVTVRAQVCDQGRVSCAKPRVDTVGLVAVGLVALLLV
LILLLLLVRRKKVKEILLPEDTRONIFYEBEGGEBEDQYIDLSQLHGLDARPE
VIRNDVAPPLMAAFQYPRFANPDEIGNFIDENIKAAADTPTAPPYDLSLVFDYEGG
SEATSLSSLNSSASDQDYDYLNEWGNRFFKLAELYGGGEDEE"
polyA_signal
polyA_site
2181..2402
ORIGIN
Alignment Scores:
Pred. No.: 7,74e-185 Length: 2436
Score: 2549.50 Matches: 480
Percent Similarity: 80.19% Conservative: 103
Best Local Similarity: 66.02% Mismatches: 137
Query Match: 58.35% Indels: 7
DB: 5 Gaps: 5

US-09-916-849A-1 (1-829) x GGCADHB (1-2436)

Qy 108 AspTrrpValValAlaProIleSerValProGluAsnGlyLysGlyProPheProGlnArg 127
Db 1 GATTGGGTGATCCCTCCCAATTAAAGTTCCTCCGAGATGAGAGGGTCCCTTCCCAAAAC 60
Qy 128 LeuAsnGlnLeuLysSerAsnLysAspArgAspThrLysIlePheTyrSerIleThrGly 147
Db 61 TTGTTTCAGATCAAAATCCAAACCGGACCGGAGCAAGATTTTCTACAGCATCACCGGG 120
Qy 148 ProGlyAlaAspSerProGluGlyValPheAlaValGluLysGluThrGlyTrrpLeu 167
Db 121 CAGGGGGCGGATGCTCCCCAGGGCATCTTACCATTGAGAGGAGAGAGCGCTGGATG 180
Qy 168 LeuLeuAsnLysProLeuAspArgGluGluIleAlaLysTyrGluLeuPheGlyHisAla 187
Db 181 AAGGTGACGACGCGCTGGACCGGAGCAGATCAACAAAGTACCACTCTACTCCCATGCC 240
Qy 188 ValSerGluAsnGlyAlaSerValGluAspProMetAsnIleSerIleValThrAsp 207
Db 241 GTGTCGAGAGATGCAACCCCGTGGAGGACCGATGAGATCATAGTGACGGTGACGGAC 300
Qy 208 GlnAsnAspHisLysProLysPheThrGlnAspThrPheArgGlySerValLeuGluGly 227
Db 301 CAAAATGACAAACAGCCCACTTCAACAGAGAGGTGTCGGGGCTCCGTGCCTGAAGGC 360
Qy 228 ValLeuProGlyThrSerValMetGlnValThrAlaThrAspGluAspAlaIleTyr 247
Db 361 GCGTGCAGGCACTTCGGTGATCGGGTGAAACGCCACCGCATGATGATGATGAGG 420

```

```

Qy 248 ThrTyrAsnGlyValValAlaTyrSerIleHisSerGlnGluProLysAspProHisAsp 267
Db 421 ACCTACACGGTGTCTATCCTCTACTCCATCCTCAGCAGGAGCGCGGAGCCCATCC 480
Qy 268 LeuMetPheThrIleHisArgSerThrGlyThrIleSerValIleSerGlyLeuAsp 287
Db 481 CACATGTTTACCGTCAACCGCGCACCGCATCCTCAGCGTCTATCGCAGCGGCTCGAC 540
Qy 288 ArgGluLysValProGluTyrThrLeuThrIleGlnAlaThrAspMetAspGly 307
Db 541 CGGGAGCGGTGCGGGAGTACACGCTGACCATGACGCGGCTGACCTGACGCGGAGGG 600
Qy 308 SerThrThrAlaValAlaValGluIleLeuAspAlaAsnAspAlaProMet 327
Db 601 CTGACCACACGCGCTGCTGTGATTGAGATCAAGACGTCAATGACAAACGCCCTGAG 660
Qy 328 PheAspProGlnLysTyrGluAlaHisValProGluAsnAlaValGlyHisGluValGln 347
Db 661 TTGATCCCAAAACGATGATGAGCGGTGCTGCGAGAGATGAGCGCGAGGTGGGC 720
Qy 348 ArgLeuThrValThrAspLeuAspAlaProAsnSerProAlaTrrpAlaThrTyrLeu 367
Db 721 CGGTTGCGCACCGACCGACCTGATGAGCGGCACACACCGCGTGGCGTCCCTCTACTCC 780
Qy 368 IleMetGlyGlyAspAspGlyAspHisPheThrIleThrThrHisProGluSerAsnGln 387
Db 781 ATCGTCCGTGGCAATGAGGCGGTGCTTACCATCATCACCCTGACCTCCAGCAACGAG 840
Qy 388 GlyIleLeuThrThrArgLysGlyLeuAspPheGluAlaLysAsnGlnHisThrLeuTyr 407
Db 841 GCGCTCTCGCACCGCCCAAGGCTGTGGACTACGAGGCCAAGCGGAGTTCGTGTCTCCAC 900
Qy 408 ValGluValThrAsnGluAlaProPheValLeuLysLeuProPheThrSerThrAlaThrIle 427
Db 901 GTGGCGGTGCTCAATGAGCTCCCTTCGCCATCAAGTCCACAGCCACGCCACCGCGTG 960
Qy 428 ValValHisValGluAspValAsnGluAlaProValPheValProSerLysValVal 447
Db 961 ATGCTCAGTGTGAGGATGTGAATGAGCGCGCTTTCGACCGCGGTGGGTGGGC 1020
Qy 448 GluValGlnGluLysLeuProThrGlyGluProValCysValTyrThrAlaGluAspPro 467
Db 1021 CAGGTCCGGAGGACGTGCGCTGGGGCAGCCCTCGCTCTCTATACAGCAGGACCT 1080
Qy 468 AspLys---GluAsnGlnLysIleSerTyrArgIleLeuArgAspProAlaGlyTrrpLeu 486
Db 1081 GACAGGCGCCAGCAGCGCATCAAGTACGTGATGGGCGAGTACCCCGCGGCTGGCTG 1140
Qy 487 AlaMetAspProAspSerGlyGlnValThrAlaValGlyThrLeuAspArgGluAspGlu 506
Db 1141 GCGGTGCACCCCGAGAACCGCTCATCAGCAGCGGAGCAGCTGGACCGTGGAG---TCC 1197
Qy 507 GlnPheValArgAsnAsnIleTyrGluValMetValLeuAlaMetAspAsnGlySerPro 526
Db 1198 CGGTTCCACCAAGACGACCTACGTGGCGGTGCTGTGGCTGTGGATGACGATGCCA 1257
Qy 527 ProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAspValAsnAspHisGlyPro 546
Db 1258 CCGCGCACAGGACCGGACCGCTGCTCTCACTGCTGATGTCAATGACCAACCGCGCT 1317
Qy 547 ValProGluProArgGlnIleThrIleCysAsnGlnSerProValArgHisValLeuAsn 566
Db 1318 GAGCGGAGCGCCCGGACATCGTCACTGTAAACCGCGCTGTGCGCGCGGTGTGAC 1377
Qy 567 IleThrAspLysAspLeuSerProHisThrSerProPheGlnAlaGlnLeuThrAsp 586
Db 1378 ATCAGCGACCGGACCTGCCCCCAACCGGGCGCTTCCGTGCGGAGTGGACCGGC 1437
Qy 587 SerAspIleTyrTrrpThrAlaGluAsnGluGlyAspThrValValLeuSerLeu 606
Db 1438 TCTGGGAGCAGCTGGGCGGTGGAGTTGCCAATGAGGAGTGACACCGTTCCTTGTGGCTG 1497

```


QY 281 ValleserSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
 Db 1076 GTGTGTCACCACTGGCTGGAGCGAGAGATTTCCTACGTATACCTGGTGGTTCAGCT 1135
 QY 301 ThrAspMetAspGlySerThrThrThrAlaValAlaValGluLeuAsp 320
 Db 1136 CTTGACCTTCAAGTGGAGGTTAAGCAACACACAGCTGTGATCAGCTCAGTGC 1195
 QY 321 AlaAsnAspAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
 Db 1196 ACCAACGATAATCTCCGATCTTCAATCCACACGTCACAGGGTCAGGTGCCTGAGAAC 1255
 QY 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
 Db 1256 GAGGCTAACGTCGTAATCACACACTGAAGTGTACTGATGTCGATGCCCCCAATACCCCA 1315
 QY 361 AlaTirpArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
 Db 1316 GCGTGGGAGGCTGTATACACCATATTG--AATGATGATGCTGGACAAATTTGCTGCACC 1372
 QY 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
 Db 1373 ACAATCCAGTGAACACGATGGCATTTTGAACACACGAAAGGCTTGGATTTTGAGGCC 1432
 QY 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
 Db 1433 AAGCAGCAGTACATCTTACAGCTAGCAGTGCAGATGTGTTGAGCTCTCTCTC 1492
 QY 421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
 Db 1493 ACCACCTCCACAGCACCGCTCCGCTGGATGTGTCGATGTGAAGGAGCCCATCTTT 1552
 QY 441 ValProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
 Db 1553 GTGCTCTGTAAGAGAGAGTGAAGTGTCCGAGCACTTTCGGCTGGCCACGAAATACACA 1612
 QY 461 ValTyrThrAlaGluAspProAspLys---GluAsnGlnLysIleSerTyrArgIleLeu 479
 Db 1613 TCCTACACTGCCAGGAGGAGACACATTTATGTAACAGAAATATACATATCGAATTGG 1672
 QY 480 ArgAspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGly 499
 Db 1673 AGAGACACTGCCAATCGCTGGAGATTATCCGACACTGTGTCACCTCCACCTCGGGCT 1732
 QY 500 ThrLeuAspArgGluAspGluGlnPheValArgAsnIleTyrGluValMetValLeu 519
 Db 1733 GAGCTGGACAGGAGGATTTGACACAGTGAAGACAGCAGTACACAGCCCTATCATATA 1792
 QY 520 AlaMetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuLeuLeu 539
 Db 1793 GCTACAGACAATGTTCTCCAGTTGCTACTGGAACAGGACACTTCTGCTGATCTGTCT 1852
 QY 540 AspValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSer 559
 Db 1853 GATGTGATGACACAGCCGCCCATACAGACCTCGAATATATTTCTCTGTGAGAGAAAT 1912
 QY 560 ProValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPhe 579
 Db 1913 CCAAAGCTCAGGTCATAAACAATCATTTGATGACAGACTTCTCCCAATACATCTCCCTTC 1972
 QY 580 GlnAlaGlnLeuThrAspAspSerAspIleTyrTyrThrAlaGluValAsnGlu---Glu 598
 Db 1973 ACAGCAGAACTAACACACGGGGCGAGTGCACCACTGGACCACTTACATACACAGCCCAACC 2032
 QY 599 GlyAspThrValValLeuSerLeuLysPheLeuLysGlnAspThrTyrAspValHis 618
 Db 2033 CAAGAATCTATATTTGAAGCCAAAGATGGCTTAGAGTGGGTGATACAAATCAAT 2092
 QY 619 LeuSerLeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCys 638
 Db 2093 CTAAGTCTAGGATAACCAAGATAAAGACCAAGTGCACCTTAGAGGTTCAGCGTGTGT 2152

QY 639 AspCysHisGlyHisValGluThrCys-----ProGlyProTyrPheGlyGlyPhe--- 655
 Db 2153 GACTGTGAAGAGGCGCGCGCGCTCTGTAGAGAGGACACAGCCTGTCAAACAGGATTGCAA 2212
 QY 656 -----IleLeuProValLeuGlyAlaValLeuAlaLeuLeuPheLeuLeuVal 672
 Db 2213 ATTCTGCGCATTTCTGGGATTTCTGGAGAAATCTTCTTCTTCTTCTTCTTCTTCTG 2272
 QY 673 LeuLeuLeuLeuValArgLysLysArgLysIleLysGluProLeuLeuLeuProGluAsp 692
 Db 2273 CTCTTCTCTTCTTCTGGAGAGAGCGGTGTCTCAAGAGCCCTTACTGCCCCCAGAGGAT 2332
 QY 693 AspThrArgAspAsnValPheTyrTyrGlyGluGluGlyGlyGlyGluLeuAsp 712
 Db 2333 GACACCGGACACACCTTTATTACTATGATGAAGAGAGGCGGAGAGAGAGAGAGAGAG 2392
 QY 713 TyrAspIleThrGlnLeuHisArgGlyLeuGluAlaArgProGluValValLeuLeuArgAsn 732
 Db 2393 TTTGACTTGCAGCAGCTGCACAGGGCTTGCAGCTCGGCTGAAGTACT---CGTAAC 2449
 QY 733 AspValAlaProThrIleIleProThrProMetTyrArgProArgProAlaAsnProAsp 752
 Db 2450 GAGTTTGCACCAACCTCATGAGTGTCCCGCTATCTTCCCGCCCTGCCAATCCCCAT 2509
 QY 753 GluIleGlyAsnPheIleIleGluAsnLeuLysAlaAlaAsnThrAspProThrAlaPro 772
 Db 2510 GAAATTTGAAATTTTATTGATGAATCTGAAGCGGCTGATCTGACCCACAGCCCGG 2569
 QY 773 ProTyrAspThrLeuLeuValPheAspTyrGluGlySerGlySerAspAlaSerLeu 792
 Db 2570 CCTTATGATTTCTGCTGCTGTTGACTATGAAGAGCGGTTCGAAAGCTCTAGTCTG 2629
 QY 793 SerSerLeuThrSerSerAlaSerAspGlnAspGlnAspTyrAspTyrLeuAsnGluTyr 812
 Db 2630 AGCTCCTCTGAATCTCTCAGCTCAGCAAGACAGGACTATGACTCTTGACGAAATGG 2689
 QY 813 GlySerArgPheLysLysLeuAlaAspMetTyrGlyGlyGlyGluAsp 829
 Db 2690 GCGAATCGCTTCAAGAGCTGCTGACATGTACGAGCGCGGCGAGCGAC 2740

RESULT 11
 HSCAD HSCAD 4778 bp mRNA linear PRI 01-JUL-1999
 LOCUS H.sapiens mRNA for E-cadherin.
 DEFINITION Z13009
 ACCESSION Z13009
 VERSION 1 GI:31072
 KEYWORDS E-cadherin.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 REFERENCE 1 (bases 1 to 4778)
 AUTHORS Bussemakers,M.J., van Bokhoven,A., Mees,S.G., Kemler,R. and Schalken,J.A.
 TITLE Molecular cloning and characterization of the human E-cadherin cDNA
 JOURNAL Mol. Biol. Rep. 17 (2), 123-128 (1993)
 MEDLINE 93211394
 PUBMED 8459805
 REFERENCE 2 (bases 1 to 4778)
 AUTHORS Bussemakers,M.J., Giroidi,L.A., van Bokhoven,A. and Schalken,J.A.
 TITLE Transcriptional regulation of the human E-cadherin gene in human prostate cancer cell lines: characterization of the human E-cadherin gene promoter
 JOURNAL Biochem. Biophys. Res. Commun. 203 (2), 1284-1290 (1994)
 MEDLINE 94380041
 PUBMED 8093045
 REFERENCE 3 (bases 1 to 4778)
 AUTHORS Bussemakers,M.
 TITLE Direct Submission
 JOURNAL Submitted (22-JUN-1992) Bussemakers M., University Hospital Nijmegen, Geert Grooteplein 16, Nijmegen, The Netherlands
 FEATURES
 1. .4778
 source


```

156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

```

```

1763 GAGCTGGACGAGGAGGATTTTGGACACGCTGAAGAACAGCACGTCACACAGCCTAATCAT 1822
520 AlaMetAspAsnGlySerProThrThrGlyThrGlyThrLeuLeuLeuLeuLeuLeu 539
1823 GCTCAGACAAATGGTCTTCCAGTCTGCTAGTGAACAGGACACTTCTCTGATCTGTCT 1882
540 AspValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSer 559
1883 GATGTGAATGACAAAGCCCATACAGAACCTCGAACTATATTCTTCTGTGAGGAGAT 1942
560 ProValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPhe 579
1943 CCAAGAGCCTCAGTCAATAACATCATTTGATGACAGACTTCTCTCCCAATACATCTCCCTTC 2002
580 GlnAlaGlnLeuThrAspAspSerAspIleTyrTyrThrAlaGluValAsnGlu--Glu 598
2003 ACAGCAGAACTAACACACGGGCGAGTGCACCTGACACCTTCAGTACAGACCCCAACC 2062
599 GlyAspThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHis 618
2063 CAAGAATCTATCATTTTGAAGCCAAAGATGGCTTAGAGTGGGTGACTACAAATCAAT 2122
619 LeuSerLeuSerAspHisGlyAsnLysGlnLeuThrValIleArgAlaThrValCys 638
2123 CTAAGCTCATGGATAACAGAAATAAGCAACAGTACACCTTAGAGGTGAGGTGCT 2182
639 AspCysHisGlyHisValGluThrCys-----ProGlyProTyrLysGlyGlyPhe--- 655
2183 GACTGTGAAGGGCGCGCGCTCTGTAGGAAGGCACAGCCTGTGGAAGCAGGATTGCAA 2242
656 -----IleLeuProValLeuGlyAlaValLeuAlaLeuLeuLeuLeuVal 672
2243 ATTCTCGCATCTCTGGGATTTCTGGAGAAATCTTGTCTTGTCTAATCTGATCTGCTG 2302
673 LeuLeuLeuValArgLysLysArgLysLysLysGluProLeuLeuLeuProGluAsp 692
2303 CTCCTGTCTTCTTCGAGGAGAGCGGTGTGTCAGAGAGCCCTTACTGCCCCAGAGGAT 2362
693 AspThrArgAspAsnValPheTyrTyrGlyGluGluGlyGlyGlyGluGluAspGlnAsp 712
2363 GACACCCGGGACACGTTTATTATCTATGATGAAGAAGAGCGGAGAGAGAGACAGGAC 2422
713 TyrAspIleThrGlnLeuHisArgGlyLeuGluAlaArgProGluValValLeuArgAsn 732
2423 TTGACTTGGAGCCACTGCACAGGCGCTGACCGCTCGGCTCGAGTACT---CGTAAC 2479
733 AspValAlaProThrIleProThrProMetTyrArgProArgProAlaAsnProAsp 752
2480 GACGTTGACCAACACCTCATGAGTGTCCCGGTATCTTCCCGCGCTGCCAATCCCGAT 2539
753 GluIleGlyAsnPheIleGluAsnLeuLysAlaAlaAsnThrAspProThrAlaPro 772
2540 GAAATGGAAATTTTATTGATGAAATCTGAAGCGGCTGATCTGACCCACAGCCCGC 2599
773 ProTyrAspThrLeuLeuValPheAspTyrGluGlySerGlySerAspAlaAlaSerLeu 792
2600 CCTTATGATCTCTGCTGCTGTTGACTATGAAGAGCGGTTCGAGAGTCTGTAGTCTG 2659
793 SerSerLeuThrSerSerAlaSerAspGlnAspGlnAspTyrAspTyrLeuAsnGluTyr 812
2660 AGCTCCCTGAACTCTCAGATCAGACAAAGACAGGACTATGACTACTTGAACGAATGG 2719
813 GlySerArgPheLysLysLeuAlaAspMetTyrGlyGlyGlyGluAspAsp 829
2720 GGCATCGCTTCAAGAACCTGGCTGACATGTACGAGGCGGCGAGGAGCAG 2770

```

RESULT 13
 HSECADH
 LOCUS
 DEFINITION
 H.sapiens mRNA for E-cadherin.
 Z18923
 ACCESSION
 VERSION
 E-cadherin.

HSECADH
 H.sapiens mRNA for E-cadherin.
 Z18923
 PRI 04-DEC-1992

[illegible]

D	b	2416	GAAATTGGAAAATTTTATGTAGTAATACTGGAAGCGGTGACTACTGACCCCAAGCCCCG	2479
Q	y	773	ProTyAspThrLeuLeuValPheAspTyrGluGlySerGlySerAspAlaAlaSerLeu	792
D	b	2476	CCTTATGATTCCTCGCTCGTGTTGTGACTATGAAGAAGCGGTTCCGAAGCTGCTAGTCTG	2535
Q	y	793	SerSerLeuThrSerSerAlaSerAspGlnAspGlnAspTyrAspTyrXleuArgGluTrp	812
D	b	2536	AGTCCCTGAACCTCCTCAGAGTCAGACAAGAACGAGGACTATGACTTCTGAACGAATGG	2595
Q	y	813	GlySerArgPhelysisysLeuAlaAspMetTyrGlyGlyGlyGluAspAsp	829
D	b	2596	GGCAATCGCTTCAAGAGCTGGCTGCATGACTGACGAGGCGCGGAGCAC	2646
RESULT 14				
ABO25106				
LOCUS Homo sapiens mRNA for E-cadherin, complete cds.				
DEFINITION ABO25106				
ACCESSION ABO25106				
VERSION ABO25106.1 GI:6682962				
KEYWORDS E-cadherin.				
SOURCE Homo sapiens (human)				
ORGANISM Homo sapiens				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eutelestomi;				
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE				
AUTHORS Shibamoto,S. and Fukudome,Y.				
TITLE E-cadherin mutant				
JOURNAL Published Only in DataBase (2000)				
REFERENCE 2 (bases 1 to 2807)				
AUTHORS Shibamoto,S. and Fukudome,Y.				
TITLE Direct Submission				
JOURNAL Submitted (19-MAR-1999) Sayumi Shibamoto, Setsunan University,				
Faculty of Pharmaceutical Sciences, Department of Biochemistry, 45-1				
Nagaotecho-cho, Hirakata, Osaka 573-0101, Japan				
(E-mail:sayumiepharm.setsunan.ac.jp, tel:81-720-66-3119,				
Fax:81-720-66-3117)				
FEATURES				
source				
1..2807				
/organism="Homo sapiens"				
/mol_type="mRNA"				
/db_xref=taxon:"9606"				
/cell_line="HSC-45M2"				
/cell_type="signet ring cell carcinoma"				
65..2770				
/codon_start=1				
/product="E-cadherin"				
/protein_id="BA88957.1"				
/db_xref="GI:6682963"				
/translation="MGFWSRKSALLLLQLVSWLQCQBEPCHPGDASSTFTVPRR HLERGLGRVNFDCTGQRATAYFSLDFTRFKVGDTGVITVKRLRFHPNQILHLYIA WSTYRKFTKYTLNTLVTHGHRRPPPHKSPYSIQAEALTFFPNSSPLGRKRQDWVIP ISCPEBNGFPFNKLVIQIKSNRPKEGFYSITQGADTFPVGVFIETRTGWLKTRA PIIDRENIKTFTLFSAHSNSGNAVEDPMELIIVTDNDNKPEFTQVEPKGSWEGL RCTSVMEVTATADDNNVTNYALAVLTLSQQPELPDKMTINRNVTGISVVTTGLD RSFSPTYLTVQNADIQQEGSLSTATATVITVDNDNPLENPTTYKGVPENEAVV ITTLTKDADAENTPAWEAIVTIINDDGQGFVVTINPVNNDGILKTAOLARCNRWHPP HCSEYCPGLGDPEAKQGYLIHVANTYPFEVSULTTSATVTVLDVNEAPIFVIPP EKREVSGFGVQEQTSTYAQBPDTFMQKIYIRWRDTANWLNLNPOTGAISFRRA LOREDTEHYKNSTYTALIATONGSPVATGTGTLIIILSDVNDNAIPFPRIIFECR NKPQVINIIDAPLNPTGPSTFAELFHGASANWTIOYNDPTQBSIILKPFMALEVGEY KINKLMNDKONDOVTVLWSCDCRGAAVCVRKAQFVBAGLCI PAILGITLGILLAL ILIILLLLFEERRRAYKVPELLPEDPTRDNVYYDEGGGEDODPDLSQLHRGLDAR PEVTRNDNVMSVPRYILPRPANPEIGNFIDENKAADDTAPTAPPDSLLVFYDEGS SSSEAASLSLNSSSESDDQDYDLNEWGNRFRKKLADMVGGGGEDD" 1202				
variation				
/note="5'base insertion between exon 9 and 10 of E-cadherin"				
ORIGIN				
Alignment Scores:				
Pred. No.: 2,18e-173 Length: 2807				
Score: 2400.50 Matches: 483				

Percent Similarity: 66.41%					
Best Local Similarity: 53.91%					
Query Match: 54.94%					
DB: 9					
Gaps: 14					
US-09-916-849A-1 (1-829) x AB025106 (1-2807)					
QY	9	AlaSerLeuLeuLeuLeuVal---	CysTrpLeuGlnCysAlaAlaSerGluProCys	27	
Dd	92	GCGCTGTCTGCCTGTCGAGTCTCTTTGGTCTC---TGCCAGGACCGGAGCCCTGC	148		
QY	28	ArgAlaValPheArgGluAlaGluValThrLeuGluAlaGlyGlyAlaGluGlnGluPro	47		
Dd	149	CACCCTGGCTTAGCGCCGAGACTACAGTTCACGGTGCCTGCCGCCCACTGGAGAGA	208		
QY	48	GlyGlnAlaLeuGlyLysVal---	PheMetGlyCysProGly----	GlnGluProAlaLeu	65
Dd	209	GGCGCGCTCTGGGACAGATGAATTTTGAAGATTGCACCGGTGCACAAGAGACACCTAT	268		
QY	66	PheSerThrAspAsnAspPheThrValArgAsnGlyGluThrValGlnGluArgArg	85		
Dd	269	TTTTTCCCTGCACACC-----CGATTCAAAGTGGGCACAGATGGTGTGATTACAGTCAAAGG	325		
QY	86	SerLeuLysGluArgAsnPro-----			92
Dd	326	CCCTACGGTTTCAATAAACCCACAGATCCATTCTTTGGTCTACGCTGGGACTCCACCTAC	385		
QY	92	-----			92
Dd	386	AGAAAGTTTCCACCAAAGTCACGCTGAATACAGTGGGGCACACACCGCCCCCGGCC	445		
QY	93	-----	LeuLysillePhePro---	SerLysArg	100
Dd	446	CATCAGSCCTCCGTTTCTGGAAATCCAAGCAGAAATGTCTCATTTTCCCACCTCTCTCCT	505		
QY	101	IleLeuArgHisLysArgAspIrpValValalaProIleSerValProGluAsnGly	120		
Dd	506	GGCCTCAGAAGACAGAAGAGACATGGGTATTCTCTCCATCACTGCCAGAAAATGAA	565		
QY	121	LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys	140		
Dd	566	AAGGCCCATTTCTAAAAACCTGGTTCAAGTCAATCCAAACAAAGACAAGAGGCAAG	625		
QY	141	IlePheTyrsertlethrGlyProGlyAlaAspSerProProGluGlyValPheAlaVal	160		
Dd	626	GTTTTCTACAGCATCACTGGCCCAAGGAGCTGACACACCCCTGTGTGTCTTTATTATT	685		
QY	161	GluLysGluThrGlyTrpLeuLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys	180		
Dd	686	GAAGAAACACAGATGGCTGAAGCTGACAGAGCCTCTGGATAGAGAACCGCATTTGCCACA	745		
QY	181	TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn	200		
Dd	746	TACACTCTTCTCTCACGCTGTCTCATCAACGGGAATGCAGTTGAGGATCCCAATGGAG	805		
QY	201	IleSertlelvalThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe	220		
Dd	806	ATTTTGATACGGTAACCGATCAGATGACAAACAGCCCGGAATTCACCCAGGAGGTCTTT	865		
QY	221	ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr	240		
Dd	866	AAGGGTCTGT CATGGAAGGTGCTCTTCCAGGAACCTCTGTGATGAGAGTACAGCCACA	925		
QY	241	AspGluAspAlaIleTyrrThrTyrsnGlyValValalatyrsertleHissertGln	260		
Dd	926	GACGCGACCATGATGTGAACACCTACAATGCCGCCCATCGCTTACACCATCTCTCAGCAA	985		
QY	261	GluProLysAspProHisAspLeuMetPheThrileHisargserThrGlyThrIleSer	280		
Dd	986	GATCCTGAGTCCCTGACAAAATAATGTTTACCATTACAGGAACACAGAGTCAATCAGT	1045		
QY	281	ValIleSerSerGlyLeuAspArgGluLysValProGluTyrrThrLeuThrIleGlnAla	300		

Db	1046	GTGCTCACCACCTGGCGCTGGACCGAGAGAGATTTCCTCAGGTATACCCCTGGTGGTTCAAGCT	1104
Qy	301	ThrAspMetAspGlyAspGlySerThrThrThralaValaValaValaGluilleLeuAsp	320
Db	1106	GCTGACCTTCAAGGTGAGGGGTTAAAGCACACACAGCAACAGCTGTGTGATCAGACTCACTGAC	1165
Qy	321	AlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn	340
Db	1166	ACCAACGATATCTCCGATCTTCAATCCCAACACAGTACAAAGGGTCAAGGTGCGCTGAGAAC	1225
Qy	341	AlaValGlyHisGluValAlaArgLeuThrValThrAspLeuAspAlaProAsnSerPro	360
Db	1226	GAGGCTAAAGCTCGTAATCACACACTGAAAGTGACTGATGCTGATGCCCCCAATACCCCA	1285
Qy	361	AlaTrpArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr	380
Db	1286	GCSTGGAGGCTGTATACACCATATTG---AATGATGATGGTGACAAATTTGTGTCACCC	1342
Qy	381	ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLys-----	394
Db	1343	ACAATCCAGTGAACACGATGGCATTTTGAAACACAGCAAAAGGACCTGGCAAGATGCAGA	1402
Qy	395	-----GlyLeuAspPheGluAlaLys	401
Db	1403	AACTGGCATCCTCACACGCTGTCATACCCCTGTCCCTGGCGCTGGATTTTGAGGCCAAG	1462
Qy	402	AsnGlnHisThrLeuTyrValGluValThrThrAsnGluAlaProPheValLeuLysLeuPro	421
Db	1463	CAGCAGTACATTCACACGTCAGTACAGCAAGTGTGTGATGTGAATGAAGCCCCCATCTTTGTG	1522
Qy	422	ThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPheVal	441
Db	1523	ACCTCCACAGCCCGTCACCGTGGATGTGTGTGATGTGAATGAAGCCCCCATCTTTGTG	1582
Qy	442	ProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCysVal	461
Db	1583	CCTCCTGAAAGAGAGTGGGAAGTGTCCGAGGACCTTGGCTGGCGCCAGGAAATCACATCC	1642
Qy	462	TyrThrAlaGluAspProAspLys---GluAsnGlnLysIleSerTyrArgIleLeuArg	480
Db	1643	TACACTGCCCGAGCCAGACACATTTATGGACACAGAAATAACATATCGAATTTGGAGA	1702
Qy	481	AspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr	500
Db	1703	GACACTGCCAACCTGGCTGGAGATTAATCCGGACACTGGTGGCCATTTCCACTCGGGGCTGAG	1762
Qy	501	LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla	520
Db	1763	CTGGACAGGGAGGATTTTGACACGCTGGAAGACAGACGTCACAGCCCTAATCATAGCT	1822
Qy	521	MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAsp	540
Db	1823	ACAGACAATGGTCTCCAGTTGTTACTTGGAAACAGGACACTTCTGTGTGATCTCTGTCTGAT	1882
Qy	541	ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro	560
Db	1883	GTGAATGACAAACGCCCCCAATACAGAACCTCGAAGCTATATTCTTCTGTGAGAGAAATCCA	1942
Qy	561	ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln	580
Db	1943	AAGCCTCAGGTCATAAACATCATTTGATGTCAGACTTCTCCCAATACATCTCCCTTCACA	2002
Qy	581	AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGlu---GluGly	599
Db	2003	GCAGAACTAAACACACGGGGCGAGTGGCCAACTGGACCATTCAGTACACAGCAACCCCA	2062
Qy	600	AspThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeu	619
Db	2063	GAATCTATCATTTTGAAGCCAAAGATGGCCCTTAGAGGTGGGTGACTACAAATCAATCTC	2122
Qy	620	SerLeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAsp	639
Db	2123	AAGCTCATGATACACAGATAAAGACCAAGTACACACCTTAGAGGTCAACGCTGTGTGAC	2182

QY 101 Ile---LeuArgAcqHisLysArgAspTrpValValAlaPheIleSerValProGluAsn 119
Db 513 CCAGGTCTCAAGAGACAGAACGAGACTGGGTCTATCCCTCCATCAGCTGCCCGGAAAT 572
QY 120 GlyLysGlyProPheProGluArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThr 139
Db 573 GAAAGGGTGAATCCCAAGAACCTGGTTCAGATCAATCCAAACAGGACCAAGAACAA 632
QY 140 LysIlePheThrSerIleThrGlyProGlyAlaAspSerProProGluGlyValPheAla 159
Db 633 AAGTTTTCTACAGCATCACCAGGCAAGAGCTGACAAACCCCGTTGGCGTTTTCATC 692
QY 160 ValGluLysGluThrGlyTrpLeuLeuAsnLysProLeuAspArgGluGluIleAla 179
Db 693 ATTGAGGGGAGACAGGCTGGCTGNAAGTGACACAGCTCTGGATAGAGAGCCATTGCC 752
QY 180 LysThrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMet 199
Db 753 AAGTACATCTCTATCTCATCGCGTGTCAATGGAAGCGGTGGAGATCCCATG 812
QY 200 AsnIleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThr 219
Db 813 GAGATAGTATCAGTACAGTACAGATCAGATGACACAGGCGCAGATTACCCAGAGGTG 872
QY 220 PheArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAla 239
Db 873 TTTGAGGATCCGTTGCAGAAAGCGCTGTTCCAGGAACCTCCGTGATGAAGTCTCAGCC 932
QY 240 ThrAspGluAspAspAlaIleThrThrAsnGlyValValAlaIleThrSerIleHisSer 259
Db 933 ACCGATCAGACGATGAGCTCAACACCTACAACTGCCATCGCTACACATCGTCAGC 992
QY 260 GlnGluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIle 279
Db 993 CAGGATCCTGAGTCCCTCACAAACATGTTCTGTCATAGGACACCGGGTTCATC 1052
QY 280 SerValIleSerSerGlyLeuAspArgGlnLysValProGluThrThrLeuThrIleGln 299
Db 1053 AGTGTCTCACTCTGGCTGGACCGAGAGATTACCTCATACATACATCTGTTGGTTCAG 1112
QY 300 AlaThrAspMetAspGlyAspGlySerThrThrThrAlaValAlaValGluIleLeu 319
Db 1113 GCTGCTGACCTTCAAGGGGAGGCTTGAGCAACACCGCCAGGCTGTGATCACTGTCAAG 1172
QY 320 AspAlaAsnAspAsnAlaProMetPheAspProGlnLysThrGluAlaHisValProGlu 339
Db 1173 GATATTAAATGACACCGTCTCTCTCAACCCGAGCAGCTATCAGGTCAGTCAAGTCCCTGAG 1232
QY 340 AsnAlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSer 359
Db 1233 AATGAGTCAATGCCCGATCGCCACACTCAAGTGACCGGATGATGATGCCCCCAACACT 1292
QY 360 ProAlaTrpArgAlaThrThrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIle 379
Db 1293 CCGCGGTGGAAGCTGTACACCGTAGTC---AACGATCCTGACACGAGCTTCGTGTGC 1349
QY 380 ThrThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGlu 399
Db 1350 GTCACAGACCCACGACCAATGATGGCATTTGAAACACAGCAAGGGCTTGGATTTTGGAG 1409
QY 400 AlalysAsnGlnHisThrLeuThrValGluValThrAsnGluAlaProPheValLeuLys 419
Db 1410 GCCAAGCAGCANTACATCTTCATGTGAGTGAGAGACAGAGNACCTTTTGGGGGTCT 1469
QY 420 LeuProThrSerThrAlaThrIleValHisValGluAspValAsnGluAlaProVal 439
Db 1470 CTTGTCCCTTCCACAGCCACTGCTGTCAGTGTGAGCGTGAATGAAGCCCCCATC 1529
QY 440 PheValProProSerLysValValGlnGluGlyIleProThrGlyGluProVal 459
Db 1530 TTTATGCTCGGAGAGAGAGTCAAGTGCAGAGTCCCGAAGACATTTGTGTGGGTCAAGAAATC 1589

QY 460 CysValThrThrAlaGluAspProAspLys---GluAsnGlnLysIleSerThrArgIle 478
Db 1590 ACATCTTATACCGCTCGAGGCGGACAGCTTCATGATCAGAAAGATCAGTATCGGATT 1649
QY 479 LeuArgAspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaVal 498
Db 1650 TGGAGGGACACTGCGCAACTGGCTGGAGATTAAACCCAGAGACTGGTCCCATTTTTCACGCGC 1709
QY 499 GlyThrLeuAspArgGluAspGluGlnPheValArgAsnAsnIleThrGluValMetVal 518
Db 1710 GCTGAGATGGACAGAGAAGACGCTGAGCATGTGAAGAACAGACATATAGTCTCTCATC 1769
QY 519 LeuAlaMetAspAsnGlySerProThrThrGlyThrGlyThrLeuLeuLeuThrLeu 538
Db 1770 ATGCGCACAGATGATGGTTCACTCCATGTCACCTGGACCGGACATCTTCTCTGGTCTG 1829
QY 539 IleAspValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGln 558
Db 1830 TTAGAGCTCAATGACACAGCTGCCATCCAGAACCTCGAAACATGCAGTCTTCCAGAGG 1889
QY 559 SerProValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerPro 578
Db 1890 AACCCACAGCTCATATCATCATCATCTTGGATCCAGACTTCCCCCAGACAGTCCCCC 1949
QY 579 PheGlnAlaGlnLeuThrAspSerAspIleThrTrpThrAlaGluValAsnGluGlu 598
Db 1950 TTTACTGCTGAGCTAACCCATGGGCGCAGCTCAACTGGACCATTTAGTATATGACGCA 2009
QY 599 Gly---AspThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrThrAspVal 617
Db 2010 GCTCAAGATCTCTCATTTTGCACCAAGAGGACTTTAGAGATTGGCGAATACAAATC 2069
QY 618 HisLeuSerLeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrVal 637
Db 2070 CATCTCAAGCTCGCGATTAACCAAGAACAGCAGGTGACACCGTGGACGTCATGTG 2129
QY 638 CysAspCysHisGlyHisValGluThrCysProGlyProThrLysGlyGlyPhe----- 655
Db 2130 TGTGACTGTGAAGGAGCGGTCAACACTGC-----ATGAAGCGCGGAATCGTGCA 2180
QY 656 -----IleLeuProValLeuGlyValValLeuAlaLeuLeuPhe 668
Db 2181 GCAGGATTCGAAGTTCCTGCCATCTCGGAATCTTGGAGGATCTCGCCCTGCTGAT 2240
QY 669 LeuLeuValValLeuLeuLeuValArgLysLysArgLysIleLysGluProLeuLeu 688
Db 2241 CTGATCCTGCTGCTCTACTGTTTCTACGAGAGAGAACGTTGTCACAAAGAGCCCTGCTG 2300
QY 689 LeuProGluAspAspThrArgAspAsnValPheThrThrGlyGluGlyGlyGly 708
Db 2301 CCACCAAGATGATACCCGGACAAATGTATATTACTATGATGAAGAGAGGAGGTGGAGAA 2360
QY 709 GluAspGlnAspThrAspIleThrGlnLeuHisArgGlyLeuGluAlaArgProGluVal 728
Db 2361 GAAGACCAAGACTTTGATTTGAGCCAGCTGCACAGGGGCTTGATGCCCGGAGAGTG 2420
QY 729 ValLeuArgAspAspValAlaProThrIleLeuProThrProMetThrArgProArgPro 748
Db 2421 ACT---CGAAATGATGTGCTCCCACTCCCTCATGAGCGTGCCTCCCTGCTCTCT 2477
QY 749 AlaAsnProAspGluIleGlyAsnPheIleIleGluAsnLeuLysAlaAlaAsnThrAsp 768
Db 2478 GCCATCTGTGTAATTTGGAACCTTCATCGATGAANACCTGAAGGACGCCACAGCGAC 2537
QY 769 ProThrAlaProProThrAspThrLeuLeuValPheAspThrGluGlySerGlySerAsp 788
Db 2538 CCCACGGCACCCCTTACGACTCTCTGTTGGTGTTCGATTTAGAGGCGAGTGGTCTCTGAA 2597
QY 789 AlaAlaSerLeuSerSerLeuThrSerSerAlaSerAspGlnAspGluAspThrAspThr 808
Db 2598 GCCGTAGCTGAGCTCACTGAATCTCTGAGTCGAGTACAGACAGACATACATATAT 2657
QY 809 LeuAsnGluTrpGlySerArgPheLysLysLeuAlaAspMetThrGlyGlyGlyGluAsp 828

Thu Sep 23 13:26:00 2004

Db 2658 CTGAACGAGTGGGCAACCGATTCAAGAGCTGGCGGACATGTACGGCGGTGGCGAGGAC 2717
Qy 829 Asp 829
Db 2718 GAC 2720

Search completed: September 23, 2004, 00:00:57
Job time : 8254 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 22, 2004, 20:38:13 ; Search time 766 seconds
(without alignments)
4597.593 Million cell updates/sec

Title: US-09-916-849a-1

Perfect score: 4369

Sequence: 1 MGUPRGPPLASLLLLQVCWLQ.....NEWGSRFKKLADNYGGGDD 829

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool.h/US09916849/runat_21092004_145802_7002/app.query.fasta_1.967
-DB=N Geneseq_29Jan04 -FASTA=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=ptc -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -NAXLEN=2000000000
-USER=US09916849 @CGN 1.1 513 @runat_21092004_145802_7002 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq_29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4369	100.0	3171	6 ABK92214	Abk92214 Prostate
2	4369	100.0	3171	6 ABN89393	Abn89393 Human P-c
3	4369	100.0	3171	6 ABT07755	Abt07755 Breast ca
4	4369	100.0	3171	6 ABK52038	Abk52038 DNA encod
5	4369	100.0	3171	6 ABN97264	Abn97264 Gene #376
6	4369	100.0	3171	7 ACAS6750	Acas6750 Human sig
7	4369	100.0	3171	7 ABZ24736	Abz24736 Human P-C
8	4369	100.0	3171	9 ADD14795	Add14795 Human src

9	4369	100.0	3777	8 ACH03801	Ach03801 Human CDN
10	4361	99.8	3205	6 Aeq81547	Aeq81547 Gene up-r
11	4361	99.8	3205	7 Aeq72821	Aeq72821 Human can
12	4361	99.8	3205	7 ABX76155	Abx76155 Lung can
13	4361	99.8	3205	7 ABX76399	Abx76399 Lung can
14	4348	99.5	3219	6 AAd27637	AAd27637 Human col
15	2449.5	56.1	4333	2 AAd11563	AAd11563 Encodes E
16	2420	55.4	4778	2 AAX56022	AAX56022 Wild-type
17	2420	55.4	4778	2 AAX56022	AAX56022 Wild-type
18	2420	55.4	4778	2 AAX56022	AAX56022 Wild-type
19	2420	55.4	4778	2 AAX56022	AAX56022 Wild-type
20	2388	54.7	2768	6 ABV94337	ABV94337 Breast ca
21	2388	54.7	2768	6 ABN89352	ABn89352 Mouse E-c
22	2388	54.7	2768	6 ABN89353	ABn89353 Mouse E-c
23	2388	54.7	2768	6 ABT42198	ABt42198 Toxicity
24	2349.5	53.8	2808	9 ADB53244	ADB53244 Primary r
25	2342.5	53.6	2808	6 AEN89392	ABn89392 Human E-c
26	2342.5	53.6	2815	2 AAd65487	AAd65487 Sequence
27	2311	52.9	2156	2 AAd11562	AAd11562 Partial e
28	1713	39.2	3427	7 ABX63589	ABx63589 Human CDN
29	1713	39.2	3427	8 ADA24495	ADA24495 Human CDN
30	1713	39.2	4122	7 ACC72793	ACC72793 Human can
31	1708	39.1	4136	5 AAS78503	AAS78503 DNA encod
32	1707	39.1	2853	6 AEN89391	ABn89391 Human N-c
33	1707	39.1	2853	7 ACC50097	ACC50097 Breast ca
34	1704.5	39.0	2824	6 AEN89350	ABn89350 Chicken N
35	1704.5	39.0	2824	6 AEN89351	ABn89351 Chicken N
36	1701.5	38.9	4350	9 ADB58603	ADB58603 Toxicity-
37	1701.5	38.9	4350	9 ADB53243	ADB53243 Primary r
38	1697.5	38.9	4321	6 AEL99593	ABs99593 Mouse isc
39	1684.5	38.6	3875	2 AAd11561	AAd11561 Partial e
40	1641	37.6	3448	9 ADD14630	ADD14630 Human src
41	1638	37.5	3310	9 ADB62391	ABd62391 Human CDN
42	1633.5	37.4	3451	9 ADD12666	Add12666 Human CDN
43	1621.5	37.1	3048	2 AAT85433	Aat85433 Human cad
44	1621.5	37.1	3048	2 AAT85433	Aat85433 Human cad
45	1621.5	37.1	3048	2 AAT61920	Aat61920 Full leng

ALIGNMENTS

RESULT 1

ABK92214

ID ABK92214 standard; DNA; 3171 BP.

XX

AC ABK92214;

XX

DT 15-AUG-2002 (first entry)

XX

DE Prostate cancer-associated DNA sequence #100.

XX

KW Prostate cancer; prostate tumour tissue; human; mammal; cytostatic;

XX Gene therapy; gene; ds.

OS Mammalia.

XX

PN WO200230268-A2.

XX

PD 18-APR-2002.

XX

PF 12-OCT-2001; 2001WO-US032045.

XX

PR 13-OCT-2000; 2000US-00687576.

PR 08-DEC-2000; 2000US-00733288.

PR 08-DEC-2000; 2000US-00733742.

PR 24-JAN-2001; 2001US-0263957F.

PR 16-MAR-2001; 2001US-0276791P.

PR 16-MAR-2001; 2001US-0276888P.

PR 06-APR-2001; 2001US-0281922P.

PR 24-APR-2001; 2001US-0286214P.

PR 30-APR-2001; 2001US-00847046.

PR 04-MAY-2001; 2001US-0288589P.

XX

QY 21 CysAlaIaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
DB 114 TGC GCGGCTCCGAGCGCGTGC GCGGCGGCTCTTCAGGAGGCTGAAGTGCCTTGGAGGCG 173
QY 41 GlyAlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
DB 174 GGAGCGCGGAGCGAGGCGCGCGCGCGCTGGGGAAATTAATCAATGGCTGCCCTGGG 233
QY 61 GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr 80
DB 234 CAAGAGCCAGCTCTGTTAGCACTGATAATGACTTCACTGTGCGGAATGGCGAGACA 293
QY 81 ValGlnGluArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
DB 294 GTCCAGGAAGAAGGTCACTGAAGAAAGGATCCATTTGAAGATCTTCCCATCCAAACGT 353
QY 101 IleLeuArgArgHisLysArgAspTrpValValAlaProIleSerValProGluAsnGly 120
DB 354 ATCTTACGACACACAGAGAGATTGGTGGTGTCTCCAAATATCTGTCCCTGAAAATGGC 413
QY 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
DB 414 AAGGGTCCCTTCCCGCAGAGACTGAATCAGCTCAAGTCTAATAAGATAGAGACCCAG 473
QY 141 IlePheTrpSerIleThrGlyProGlyAlaAspSerProGluGlyValPheAlaVal 160
DB 474 ATTTTACACGATCACGGGCGCGGGGCGAGACAGCCCGCTCAGGGGTGTCTTCGCTGTA 533
QY 161 GluLysGluThrGlyTrpLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys 180
DB 534 GAGAAGGAGACAGGCTGTTGTTGTTGAATGAAGCCACTGGACCGGGAGGAGATTGCCAAG 593
QY 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
DB 594 TATGAGCTTTTGGCCACGCTGTGTACAGNATGGTGCCTCAGTGGAGGACCCCATGAAC 653
QY 201 IleSerIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
DB 654 ATCTCCATCATCGTACCGACAGCAAGTACACCAAGCCCAAGTTTACCCAGGACACCTTC 713
QY 221 ArgGlySerValLeuGluGlyValIleuProGlyThrSerValMetGlnValThrAlaThr 240
DB 714 CGAGGAGTGCTTAGAGGAGTCTACCGAGTACTTCTGTGATGACAGGTGACAGCCACA 773
QY 241 AspGluAspAspAlaIleThrThrAsnGlyValValAlaTrpSerIleHisSerGln 260
DB 774 GATGAGGATGATGCCATCTACACTACAAATGGGGTGGTGTGTTACTCCATCCATAGCCAA 833
QY 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
DB 834 GAACCAAGAGACCCACAGCCTCATGTTCACATTCACCGAGCAGGACCCATCAGC 893
QY 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTrpThrLeuThrIleGlnAla 300
DB 894 GTCATCTCCAGTGGCTGACCGGGAAAAAGTCCCTGAGTACACACTGACCATCCAGGCC 953
QY 301 ThrAspMetAspGlyAspGlySerThrThrThrAlaValAlaValGluIleLeuAsp 320
DB 954 ACAGACATGGATGGGAGCGGTCCACCAACCGCAGTGGCAGTAGTGGAGATCCCTTGTAT 1013
QY 321 AlaAsnAspAsnAlaProMetPheAspProGlnLysTrpGluAlaHisValProGluAsn 340
DB 1014 GCCAATGCAATGTCTCCATGTTTGAACCCCAAGATACGAGGCCCATGTGCTGAGAAT 1073
QY 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
DB 1074 GCAGTGGGCCATGAGTGCAGAGGCTGACGGTCACTGTATCTGGACGCGCCCACTCACCA 1133
QY 361 AlaTrpArgAlaThrTrpLeuIleMetGlyValAspAspGlyAspHisPheThrIleThr 380
DB 1134 GCGTGGCGTGCACCTACTCTATCATGGCGCGGTGACGCGGGGACCAATTTTACCATCACC 1193
QY 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400

DB 1194 ACCCACTCTGAGAGCAACAGGGCATCTGACNACAGNAGGTTTGGATTTTGGAGGCC 1253
QY 401 LysAsnGlnHisThrLeuTrpValGluValThrAsnGluAlaProPheValLeuLysIleu 420
DB 1254 AAAAACCCAGCACACCCTGTAGCTTGAAGTACCACAGAGGCCCTTTTGTCTGAAGCTC 1313
QY 421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
DB 1314 CCAACTTCAAGCCACCATAGTGTCTCCAGTGGAGATGTGAATGAGGCACCTGTGTTT 1373
QY 441 ValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
DB 1374 GTCCCACTCCAAAGTCTGTTGAGTCCAGAGGAGCATCCCACTGGGAGGCTGTGTGT 1433
QY 461 ValTrpThrAlaGluAspProAspLysGluAsnGlnLysIleSerTrpArgIleLeuArg 480
DB 1434 GTCTACACTGCAGAGACCTTGACAAAGAGATCAAAAGATCAGTACCCTCCGATCTCTAGA 1493
QY 481 AspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
DB 1494 GACCCAGCAGGGTGGCTAGCCATGACCCAGACAGTGGGCAAGTCAACGTGTGGGACC 1553
QY 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTrpGluValMetValLeuAla 520
DB 1554 CTGCACTGTAGGATGAGCAGTGTGTGAGGAAACAACATCTATGAAGTCACTGGTCTGGCC 1613
QY 521 MetAspAsnGlySerProProThrThrGlyThrLeuLeuLeuThrLeuIleAsp 540
DB 1614 ATGGCAATGGAGCCCTCCCACTGGCAGCGAACCCTTCTGCTAACACTGATTGAT 1673
QY 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
DB 1674 GTCAAGACCATGGCCAGTCCCTGAGCCCGTGCAGATCACCATCTGCAACAAGCCCT 1733
QY 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
DB 1734 GTGCGCCAGTGTGAACATCAGCAAGGACCTGTCTCCCAACACTCCCTTTCAG 1793
QY 581 AlaGlnLeuThrAspAspSerAspIleTrpThrAlaGluValAsnGluGluGlyAsp 600
DB 1794 GCCAGCTCACAGATGACTCAGACATCTACTGACCGCAGAGGTCAACAGAAAGGTGAC 1853
QY 601 ThrValValLeuSerIleLysPheLeuLysGlnAspThrTrpAspValHisLeuSer 620
DB 1854 ACAGTGGTGTCTGCTGAAGAGTTCCTGAAGCAGGATACATATGAGCTGCACCTTCT 1913
QY 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
DB 1914 CTGCTGACCATGGCAACAAGAGCAGCTGACGGTGCATCAGGGCCACTGTGTGGACTGC 1973
QY 641 HisGlyHisValGluThrCysProGlyProTrpLysGlyGlyPheIleLeuProValLeu 660
DB 1974 CATGGCCATGTGAAACCTGCCCTCGAACCTGAAAGAGGTTTCATCCTCCCTGTGTG 2033
QY 661 GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuValArgLysLys 680
DB 2034 GGGGCTGCTCTGCTCTGCTGTCTCTCTGCTGGTGTGCTTGTGTGTGGAGAAAG 2093
QY 681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTrp 700
DB 2094 CGGAAGATCAAGGAGCCCTCTTACTCCAGAAAGATGACCCGCTGACACAGCTCTTCTAC 2153
QY 701 TrpGlyGluGluGlyGlyGlyGluAspGlnAspTrpAspIleThrGlnLeuHisArg 720
DB 2154 TATGGCAAGAGGGGGTGGCGAAGAGACCAAGGACTATGACATCACCACCTCCACCA 2213
QY 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
DB 2214 GTCTGGAGGCCAGCGCGAGGTGTCTCTCGCAATGACGTGGCACCACCAACCATCTCCG 2273
QY 741 ThrProMetTrpArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu

Db 894 GTCATCTCCAGTGGCTGGACCGGAAAAAGTCCCTGAGTACACACTGACCATCCAGGCC 953
 Qy 301 ThrAspMetAspGlyAspGlySerThrThrAlaValAlaValGluLeuAsp 320
 Db 954 ACAGACATGGATGGGAGCGGCTCCACACACAGCGGAGTGGAGATCCTTGAT 1013
 Qy 321 AlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
 Db 1014 GCCAATGACAAATGCTCCCATGTTTACCCCGAAGTACGAGGCCCATGTCCTGAGAT 1073
 Qy 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
 Db 1074 GCAGTGGGCCATGAGTGCAGAGGCTGACGGTCACTGATCTGGACGCCCCCACTACCA 1133
 Qy 361 AlaTrpArgAlaThrTyrLeuIleMetGlyAspAspGlyAspHisPheThrIleThr 380
 Db 1134 GCGTGGCTGGCCACTACCTTATCATGGCGGTGACGACGGGACCACTTTTACCATCACC 1193
 Qy 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
 Db 1194 ACCACCCCTGAGAGCAACAGGGCATCTGACACACAGGAGGTTTGGATTTGAGGCC 1253
 Qy 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGlnAlaProPheValLeuLysLeu 420
 Db 1254 AAAAACACGACACCTGTACCTTGAAGTGAACACAGGCGCCCTTTTGTGCTGAAGCTC 1313
 Qy 421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
 Db 1314 CCAACCTCCACAGCCACCATGAGTGTGTCAGTGGAGGATGTGATGAGGACCTGTGTTT 1373
 Qy 441 ValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
 Db 1374 GTCCACCCCTCAAAGTGGTGGTCCAGAGGGCATCCCACTGGGAGCCCTGTGTGT 1433
 Qy 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
 Db 1434 GTCTACACTGAGAGAGCCCTGACAGAGGAGATCAAAAGATCAGTACCGCATCCTGAGA 1493
 Qy 481 AspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
 Db 1494 GACCCAGCAGGGTGGCTAGCCATCGACACAGCAGTGGGAGGTCACAGCTGTGGGCACC 1553
 Qy 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
 Db 1554 CTCGACCTGAGGATGAGCATGTTGTGAGGAGCAACATCTATGAAGTCATGCTTGGCC 1613
 Qy 521 MetAspAsnGlySerProProThrThrGlyThrLeuLeuLeuLeuThrLeuIleAsp 540
 Db 1614 ATGGACATGGAGCCCTCCACACCTGGCAGCGGACCCCTTCTGCTAACACTGATTGAT 1673
 Qy 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
 Db 1674 GTCAACGACCATGGCCAGTCCCTGAGCCCGTCAGATCAACATCTGCAACCAAGCCCT 1733
 Qy 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
 Db 1734 GTGCCACATGGATGAAATCAGCAGCAAGAGACTGTCTCCACACCTCCCTTCCAG 1793
 Qy 581 AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGlyAsp 600
 Db 1794 GCCCAGCTCAGATGACTCAGACATCTACTGAGCGGAGAGGTCACAGGAGAGGTGAC 1853
 Qy 601 ThrValValLeuSerLeuLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
 Db 1854 ACAGTGGTCTTGTCCCTGAAAGAGTCTTGAAGCAGGATACATATGACGTGACCTTCT 1913
 Qy 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
 Db 1914 CTGCTGACCATGGCAACAAAGAGCAGTGCAGCGTGCATCAGGCGCATCTGTGTGCGACTG 1973
 Qy 641 HisGlyHisValGluThrCysProGlyProTyrLysGlyPheIleLeuProValLeu 660
 Db 1974 CATGGCCATGTGAAACCTGGACCTGGAAAGAGGAGTTCCTCCTCCCTGTGCTG 2033

Qy 661 GlyAlaValLeuAlaLeuPheLeuLeuValLeuLeuValLeuValArgLysLys 680
 Db 2034 GGGGCTGCTCCTGGCTCTGCTGCTCTCTGCTGGTGGTCTGCTTTGTTGGTGAAGAAAG 2093
 Qy 681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
 Db 2094 CGGAAGATCAAGGAGCCCTCTACTCCAGAGATGACACCCGTGACACGCTCTCTAC 2153
 Qy 701 TyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
 Db 2154 TATGCCAAGAGCGGGGTGGCGAAGAGACCAAGGACTATGATCACCAGCTCCACCGA 2213
 Qy 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
 Db 2214 GGTCTGGAGCCAGCGCGGAGGTGTTCTCCGANTGACGTGGCACCACCATCATCCCG 2273
 Qy 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
 Db 2274 ACACCCATGTACCGTCTAGGCGACCAACCCAGATGAATCGGCAACTTTATATTGAG 2333
 Qy 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
 Db 2334 AACCTGAAGCGGCTTACACAGACCCACAGCCCGCCCTACGACACCTCTTGGTGTTC 2393
 Qy 781 AspTyrGluGlySerGlySerAspAlaSerLeuSerSerLeuThrSerSerAlaSer 800
 Db 2394 GACTATGAGGCGAGCGGCTCCGACCGCGTCCCTGAGTCCCTCACCTCCCTCCGCTCC 2453
 Qy 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTyrGlySerArgPheLysLeuAla 820
 Db 2454 GACCAAGACCAAGATTACGATTATCTGAACGAGTGGGCGAGCGCTTCAAGAGCTGSCA 2513
 Qy 821 AspMetTyrGlyGlyGlyGluAspAsp 829
 Db 2514 GACATGACGCTGGCGGGGAGGACGAC 2540
 RESULT 4
 ABK52038
 ID ABK52038 standard; DNA; 3171 BP.
 XX
 AC ABK52038;
 XX
 DT 13-AUG-2002 (first entry)
 XX
 DE DNA encoding human P-cadherin (placental cadherin) protein.
 XX
 KW Human; P-cadherin; placental cadherin; solid surface; vascular graft;
 cell-to-cell cohesion; endothelial cell-coated surface;
 KW hydrodynamic shear; shaking container; continuous harvest system;
 KW laminar shear device; gene; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 PH 54..2543
 CDS /tag= a
 FT /product= "Human placental cadherin (P-cadherin) protein"
 FT polyA_signal 3162..3167
 FT /tag= b
 FT /standard_name= "PolyA signal"
 XX
 XX WO200231121-A2.
 PN
 XX
 XX 18-APR-2002.
 PD
 XX
 XX 15-OCT-2001; 2001WO-US032030.
 PF
 XX
 XX 13-OCT-2000; 2000US-0241216P.
 PR
 XX 27-OCT-2000; 2000US-0243693P.
 PR
 XX 11-OCT-2001; 2001US-00975723.
 PR
 XX (UTNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.
 PA

Db 1674 GTCAACGACCATGCGCCAGTCCCTGAGCCGTCAGATCACCATCTCCAAACCAAGCCCT 1733
 Qy 561 ValArgHisValLeuAenIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
 Db 1734 GTGGCCACGCTGCTGAACATCACGACCAAGACCTGTCTCCCCACACCTCCCTTTCCAG 1793
 Qy 581 AlaGlnLeuThrAspAspSerAspIleTyrTrpAlaGluValAsnGluGluGlyAsp 600
 Db 1794 CCCAGCTCACAGATGACTCAGATCATCTAGTGCAGCGCAGAGGTCACAGAGGAGTGAC 1853
 Qy 601 ThrValValLeuSerLeuLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
 Db 1854 ACAGTGGTCTTGTGCTCCGCTGAAGAGTTCCTGAAGCAGGATACATATGACGTGCACCTTTCT 1913
 Qy 621 LeuSerAspHisGlyAsnLysGluLeuThrValIleArgAlaThrValCysAspCys 640
 Db 1914 CTGTCTGACCATGGCAACAAAGACAGCTGACGGTGTATCAGGGCCATGTGTGGACTGC 1973
 Qy 641 HisGlyHisValGluThrCysProGlyProTyrPheGlyGlyPheIleLeuProValLeu 660
 Db 1974 CATGCCCATGTGCAAAACCTGCGCTGGACCTCGGAAGAGGTTTCACTCTCCCTGTGCTG 2033
 Qy 661 GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuValLeuValArgLysLys 680
 Db 2034 GGGGCTGTCTGGCTGCTGCT 2093
 Qy 681 ArgLysLysGluProLeuLeuLeuProGluAspAspThrArgAspAspValPheTyr 700
 Db 2094 CGGAAGATCAAGAGGCCCT 2153
 Qy 701 TyrGlyGluGluGlyGlyGlyGluAspGlnAspThrAspIleThrGlnLeuHisArg 720
 Db 2154 TATGCGAAGAGGGGGTGGGGAAGAGGACCAAGCACTATGACATCACCAGCTCCACCGA 2213
 Qy 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
 Db 2214 GGTCTGGAGGCGAGCGGAGTGTCTCTCGCAATGACGTGGCACCACCATCATCCCG 2273
 Qy 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
 Db 2274 ACACCCATGTACCGTCTTAGCCAGCCAAACCCAGATGAAATCGGCACACTTATAATAG 2333
 Qy 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuValPhe 780
 Db 2334 AACCTGAAGGCGGCTTAAACACAGACCCCAAGCCCGCCCTACGACACCCCTCTTGGTTC 2393
 Qy 781 AspTyrGluGlySerGlySerAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
 Db 2394 GACTATGAGGCGAGCGGCTCCGAGCGCGTCCCTGAGCTCCCTCAGCTCCCTCGCCCTCC 2453
 Qy 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluThrGlySerArgPheLysLysLeuAla 820
 Db 2454 GACCAAGACCAAGATTACGATTATCTGAACAGTGGGCGAGCCGCTTCAAGAGCTGGCA 2513
 Qy 821 AspMetTyrGlyGlyGluAspAsp 829
 Db 2514 GACATGTACGTTGGCGGGGAGGAGCAGC 2540
 RESULT 5
 ID ABN97264
 XX ABN97264 standard; DNA; 3171 BP.
 AC ABN97264;
 XX
 DT 13-AUG-2002 (first entry)
 DE Gene #3762 used to diagnose liver cancer.
 DE
 KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
 KW metastatic liver tumor; cytostatic; expression profile; disease state;
 KW disease progression; drug toxicity; drug efficacy; drug metabolism.
 XX Homo sapiens.
 OS

XX PN WO200229103-A2.
 XX PD 11-APR-2002.
 XX PF 02-OCT-2001; 2001WO-US030589.
 XX PR 02-OCT-2000; 2000US-0237054P.
 XX PA (GENE-) GENE LOGIC INC.
 XX PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
 XX MPI; 2002-426119/45.
 XX DR
 XX PT Diagnosing and detecting the progression of liver cancer, hepatocellular
 PT carcinoma or metastatic liver tumor in a patient, involves detecting the
 PT level of expression of two or more genes in a liver tissue sample.
 XX
 PS Claim 1; SEQ ID NO 3762; 298pp; English.
 XX
 CC The invention relates to a novel method for diagnosing and detecting the
 CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
 CC tumor in a patient, and differentiating metastatic liver cancer from
 CC hepatocellular carcinoma in a patient, involving detecting the level of
 CC expression of two or more genes represented in ABN93503-ABN97455 in a
 CC tissue sample. The method of the invention has hepatotropic, and
 CC cytostatic activity. The method is useful for diagnosing and detecting
 CC the progression of liver cancer, hepatocellular carcinoma and metastatic
 CC liver carcinoma in a patient. The method is useful for identifying
 CC expression profiles which serve as useful diagnostic markers as well as
 CC markers that can be used to monitor disease states, disease progression,
 CC drug toxicity, drug efficacy and drug metabolism. Note: The sequence data
 CC for this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 3171 BP; 740 A; 903 C; 864 G; 664 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 0 Length: 3171
 Score: 4369.00 Matches: 829
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-09-916-849A-1 (1-829) x ABN97264 (1-3171)

Qy 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuLeuValCysTrpLeuGln 20
 Db 54 ATGGGGCTCCCTCGTGGACCTCTCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 113
 Qy 21 CysAlaAspSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
 Db 114 TCGCGCGCTCCGAGCGGCTCGCGCGCTCTTCAGGAGGCTGAAAGTACCTTGAGGCGG 173
 Qy 41 GlyGlyValaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
 Db 174 GGAGCGCGGAGCAGAGCGCCGCGCGCTGGGAAAGTATTTCATGGGCTCGCCCTGGG 233
 Qy 61 GlnGluProAlaLeuPheSerThrAspAsnAspAspPheThrValArgAsnGlyCysThr 80
 Db 234 CAAGAGCCAGCTCTGTGTTAGCACTGATAATGATGACTTCACTGTCCGGAATGGCGAGACA 293
 Qy 81 ValGlnGluArgArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
 Db 294 GTCCAGAAAGAGGTCACTGAGGAAAGAAATCCATTGAAGATCTTCCCATCCAAAGCT 353
 Qy 101 IleLeuArgArgHisLysArgAspTrpValValAlaProIleSerValProGluAsnGly 120
 Db 354 ATCTTACGAAGACACAAAGAGAGATTGGGTGGTGTCTCCAAATATCTGTCTCCGAAAAATGCG 413

QY 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
Db 414 AAGGTCCTCCCTCCCAAGAGAGCTGAATCAGCTCAAGTCTAATAAAGATAGAGACACCAAG 473
QY 141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProGluGlyValPheAlaVal 160
Db 474 ATTTTCTACAGCATCAGGGGCGGGGCGAGACAGCCCCCTGAGGGTGTCTTCGCTGTA 533
QY 161 GlutysGluThrGlyTrpLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys 180
Db 534 GAGAAGGAGAGCAGGCTGTGTGTGTTGAATAAGCCACTGGACGGGAGGAGATTGCCAAG 593
QY 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
Db 594 TATGAGTCTTTGGCCACGCTGTGTGAGAGATGGTGCCTCAGTGGAGGACCCCATGAAC 653
QY 201 IleSerIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
Db 654 ATCTCCATCATCGTACCGCAGCATGACCAAGCCCAAGTTTACCCAGGACACCTTC 713
QY 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
Db 714 CGAGGGAGTGTCTAGAGGAGTCCCTACCAAGTACTTCTGTGATGCGAGGTGAGCCACA 773
QY 241 AspGluAspAspAlaIleTyrThrTyrAsnGlyValAlaTyrSerIleHisSerGln 260
Db 774 GATGAGGATGATGCATCTACACCTACATCAATGGGGTGTGTCTTACTTCCATCCATAGCAA 833
QY 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
Db 834 GAACCAAGAGCCACACGACCTCATGTTCCAAATTCACCGGAGCAGGACACCATCAGC 893
QY 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
Db 894 GTCATCTCCAGTGGCTGGACCGGGAAGTCCCTGAGTACACACTGACCATCCAGGCC 953
QY 301 ThrAspMetAspGlyAspGlySerThrThrAlaValAlaValGluIleLeuAsp 320
Db 954 ACAGACATGATGGGACGGCTCCACACACGCGAGTGGCAGTGGAGATCTTTGAT 1013
QY 321 AlaAsnAspAlaProMetPheAspProGluLysTyrGluAlaHisValProGluAsn 340
Db 1014 GCCATGACATGCTCCATGTTTGACCCCAAGTACGAGGCCCATGTGCTTGAGAT 1073
QY 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
Db 1074 GCACTGGGCCATGAGGTGACAGGCTGACGCTCACTGATCTGGACGCCCCCACTCACCA 1133
QY 361 AlaTrpArgAlaThrTyrLeuIleMetGlyAspAspGlyAspHisPheThrIleThr 380
Db 1134 GCGTGGCGTCCACCTACCTTATCANGGCGGTGACGACGGGACCATTTTACCATCAC 1193
QY 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
Db 1194 ACCACCTTGAGACACACCGGCGATCTCGACACCGAGGGTTTGGATTTTGAGGCC 1253
QY 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db 1254 AAAAAACGACACACCTGTACGTTGAAGTACCAACGAGGCCCTTTTGTGCTGAAGCTC 1313
QY 421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
Db 1314 CCAACCTCCACAGCCACATAGTGTCCACGTGAGGATGTGAATGAGGCACCTGTGTTT 1373
QY 441 ValProSerLysValValGluValGlnGluIleProThrGlyGluProValCys 460
Db 1374 GTCCACCTCCAAAGTCGTTGAGTCCAGAGGGCATCCCCACTGGGAGCCTGTGTGT 1433
QY 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db 1434 GTCTACACTGCAGAGACCTCGACAGGGAATCAAAAGATCAGCTACCGCATCCTCGAGA 1493
QY 481 AspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500

Db 1494 GACCCAGCAGGCTGGCTAGCCATGACCCAGACAGTGGGCAGGTACAGCTGTGGGCACC 1553
QY 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db 1554 CTCGACCGTGAAGATGAGCAGTTTGTGAGGAACAACATCTATGAAGTCAATGCTGTGGCC 1613
QY 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAsp 540
Db 1614 ATGGCAATGGGAAGCCCTCCCACTGGACACCGGAACCTTCTGCTTAACACTGATTGAT 1673
QY 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 1674 GTCACAGCATTGGCCAGTCCCTGAGCCCCGTGAGATCACCATCTGCAACCAAGCCCT 1733
QY 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
Db 1734 GTGCGCCACGCTGTAACATCACGGAACAGGACCTGTCTCCCAACACCTCCCTTTCCAG 1793
QY 581 AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGluGlyAsp 600
Db 1794 GCCCAGCTCACAGATGATCTCAGACATCTACTGGACGGCAGAGGTCAACGAGGAAGTAC 1853
QY 601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
Db 1854 ACAGTGTCTTGTCTGTAAGAAGTTCCTGAAGCAGGATACATATGACGTGCACCTTCT 1913
QY 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
Db 1914 CTGTCTGACCATGGCAACAAGAGCAGCTGACGGTGATCAGGGCCACTGTGTGCACCTGC 1973
QY 641 HisGlyHisValGluThrCysProGlyProTyrLysGlyGlyPheIleLeuProValLeu 660
Db 1974 CATGGCCATGTCGAACCTGCCCTGGAACCTGGAAAGAGGTTTCATCTCCTCCTGCTG 2033
QY 661 GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysLys 680
Db 2034 GGGGCTGTCTGCTGTCTGCTGTCTGCTGTCTGCTGTCTGCTGTCTGCTGTCTGCTG 2093
QY 681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
Db 2094 CGAGATCAAGAGGCCCTCTCTACTCCAGNAGATGACCCCGTGACACCTCTTCTAC 2153
QY 701 TyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
Db 2154 TATGGCAAGAGGGGGTGGGAGAGGACCAAGGACTATGACATCACCCAGCTCCACCGA 2213
QY 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleLeuPro 740
Db 2214 GGTCTGGAGGCGCAGGCGGAGTGTCTTCGCAATGACGTGGCACCACCATCATCCCG 2273
QY 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
Db 2274 ACACCCATGTACCGTCTTAGGCCACCAACCCAGATGAAATCGGCACTTTATATTGAG 2333
QY 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
Db 2334 AACCTGAAGGGCGCTTAACACAGACCCCGCCCTACGACACCTCTTGTGTGTC 2393
QY 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
Db 2394 GACTATGAGGCGAGGCTCCGACCGCGCTCCCTGAGCTCCCTCACCTCTCCGCTCC 2453
QY 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGlySerArgPheLysLeuAla 820
Db 2454 GACCAAGACCAAGATTATCTGAACGAGTGGGCGAGCCGCTTCAAGAAAGTGGCA 2513
QY 821 AspMetTyrGlyGlyGluAspAsp 829
Db 2514 GACATGTACGTTGGCGGAGAGGACGAC 2540

RESULT 6
ACA56750

ID	ACA56750 standard; cDNA; 3171 BP.	21	CysAlaIaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla	40
XX				
AC	ACA56750;	114	TGCGCGGCTCGAGCGCTGCGGCGGTCTTCAGGGAGGCTGAAGTACCTTGGAGCG	173
XX				
DT	06-JUN-2003 (first entry)	41	GlyGlyValaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly	60
XX				
DE	Human signalling pathway polynucleotide probe SEQ ID NO 1348.	174	GGAGCGGGAGCAGGAGCGCGCGCAGCGCTGGGAAAGTATTTCATGGGCTGCCCTGGG	233
XX				
KW	Human; probe; ss; array element; Parkinson's disease;	61	GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr	80
KW	signalling pathway population; cancer; adenocarcinoma; leukaemia;	234	CAAGAGCCAGCTCTCTTTAGCACTGATAATGATGACTTCACTGTCCGGAATGGCGACA	293
XX	immunopathy; AIDS; asthma; neuropathy; Alzheimer's disease; microarray.			
XX		81	ValGlnGluArgArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg	100
OS	Homo sapiens.	294	GTCCAGGAAGAAGGTCACTGAGGAAGAAGATCCATTGAAGATCTTCCATCCAAAGCT	353
XX				
FN	US6500938-B1.	101	IleLeuArgArgHisLysArgAspTrpValValaProIleSerValProGluAsnGly	120
XX				
PD	31-DEC-2002.	354	ATCTTACGAAGACACAGAGAGATTGGGTGTTGCTCCATATCTGTCCTGAAATGGC	413
XX				
PF	30-JAN-1998; 98US-00016434.	121	LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspAspThrLys	140
XX				
PR	30-JAN-1998; 98US-00016434.	414	AAGGGTCCCTTCCCCAGAGACTGAATCAGCTCAAGTCTAATAAAGATAGAGACCCAG	473
XX	(INCY-) INCYTE GENOMICS INC.			
PA		141	IlePheTyrSerIleThrGlyProGlyAlaAspSerProProGluGlyValPheAlaVal	160
XX				
PI	Au-Young J, Seilhamer JJ;	474	ATTTTCTACAGCATCACGGGGCGGGGCGAGACAGCCCCCTGAGGGTGTCTTCGTGTA	533
XX				
DR	WPI; 2003-352189/33.	161	GluLysGluThrGlyTrpLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys	180
XX				
PT	Combination of polynucleotide probes, useful as array elements in a	534	GAGAAGGAGACAGGGTGGTGTGTTGTTGAATAAGCCACTGGACCGGGAGGAGATTGCCAAG	593
PT	microarray for monitoring the expression of a number of target			
PT	polynucleotides.			
XX		181	TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn	200
PS	Claim 1; SEQ ID NO 1348; 65pp; English.	594	TAAGAGCTCTTTGGCCAGCTGTGTGTCAGAGATGGTGCCTCAGTGGAGGCCCATGAAC	653
XX				
CC	The invention relates to a combination which, comprises a number of	201	IleSerIlelleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe	220
CC	polynucleotide probes comprising a sequence selected from one of the 1490			
CC	sequences mentioned in the specification. The combination is useful as an	654	ATCTCCATCATGTCGACCGACAGCAATGACCACAGCCCAAGTTTACCAGGACACCTTC	713
CC	array element in a microarray for monitoring the expression of a number			
CC	of target polynucleotides. The microarray is particularly useful in the			
CC	diagnosis and treatment of cancer and immunopathology and neuropathology.	221	ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr	240
CC	The microarray is useful in diagnostics and treatment regimens, drug	714	CGAGGAGTGCTTAGAGGGAGTCTCTACAGGTACTTCTGTGATGCAGGTGCACAGCCACA	773
CC	discovery and development, toxicological and carcinogenicity studies,			
CC	forensics and pharmacogenomics. The microarray is also useful for	241	AspGluAspAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln	260
CC	monitoring progression of diseases and for developing sophisticated			
CC	profiles for the effects of currently available therapeutic drugs. The	834	GAACCAAGGAGCCACACAGCTCATGTTCACAAATCCAGGAGGACAGGACCATCAGC	893
CC	combination is also useful for purifying a subpopulation of mRNAs, cDNAs			
CC	and genomic fragments and in research and diagnostic applications. The	774	GATGAGGATGATGCCATCTACACCTACAAATGGGTGGTGTCTTACTCCATCCATAGCCAA	833
CC	array can detect changes in expression in a large number of genes coding			
CC	for different signaling pathway populations which can be used to diagnose	261	GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer	280
CC	various diseases including cancer e.g. adenocarcinoma and leukaemia,			
CC	immunopathies e.g. AIDS and asthma, neuropathies e.g. Alzheimer's disease	894	GTCACTCCAGTGGCTTGACCGGGAAGTCCCTGAGTACACACTGACCATCCAGGCC	953
CC	and Parkinson's disease. The present sequence represents a polynucleotide			
CC	probe of the invention. Note: The sequence data for this patent did not	301	ThrAspMetAspGlyAspGlySerThrThrThrAlaValAlaValGluIleLeuAsp	320
CC	form part of the printed specification but was obtained in electronic			
CC	format directly from USPTO at	954	ACAGACATGGATGGGAGCGGTCTCCACCCAGCGAGTGGCAGTAGTGGATCTCTTGAT	1013
CC	seqdata.uspto.gov/sequence.html?DocID=06500938B1			
XX		321	AlaAsnAspAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn	340
XX				
Seq	Sequence 3171 BP; 740 A; 903 C; 864 G; 664 T; 0 U; 0 Other;	1014	GCAATGACATGCTCCCATGTTTGACCCCCCAGAGTACGAGGCCCATGTGCTGAGAAAT	1073
Alignment Scores:				
Pred. No.:	0	Length:	3171	
Score:	4369.00	Matches:	829	
Percent Similarity:	100.00%	Conservative:	0	
Best Local Similarity:	100.00%	Mismatches:	0	
Query Match:	100.00%	Indels:	0	
DB:	7	Gaps:	0	
US-09-916-849A-1 (1-829) x ACA56750 (1-3171)				
Qy	1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuGlnValCysTrpLeuGln	20		
Db	54 ATGGGCTCCCTCGTGGACCTCTCGCGTCTCTCCTCTCTCCAGGTTTGTGGCTGCAG	113		

Db 1194 ACCCACCCTGAGACCAACCGGGCATCTGACACACCGAGGAGGGTTGGATTGTGAGGCC 1253
Qy 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db 1254 AAAAAACGAGCACACCCCTGTAGTTGAAGTGAACCAACGAGGCCCTTTTGTGCTGAAGCTC 1313
Qy 421 ProThrSerThrAlaThrIleValHisValGluAspValAsnGluAlaProValPhe 440
Db 1314 CCAACCTCCACAGCCACCATAGTGTCCAGTGGAGGATGTGAATGAGGACACCTGTGTTT 1373
Qy 441 ValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
Db 1374 GTCCACCCCTCCAAAGTCGTGTAGTCCAGAGGGCATCCCACTGGGAGGCTGTGTGT 1433
Qy 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db 1434 GTCTACACTGTCAGAGAACCCTGACAGAGAGAAATCAAAAGATCAGCTACCGCATCTGAGA 1493
Qy 481 AspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db 1494 GACCCAGAGGTGGCTAGCCATGGACCCAGACAGTGGGAGGTACAGCTGTGGGACC 1553
Qy 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db 1554 CTCGACCGTGAGGATGAGCAGTTTGTGAGGAACAACATCTATGAAGTCATGGTCTTGGCC 1613
Qy 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuLeuAsp 540
Db 1614 ATGGACAATGGAAGCCCTCCACCATCGGCACGCGGAACCCCTTCTGTGAACACTGATTAT 1673
Qy 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 1674 GTCAACGACCATGGCCAGTCCCTGAGCCCGTCCAGATCACCATCTGCAACCAAGCCCT 1733
Qy 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
Db 1734 GTGGCCACGCTGTGAACATACGACGACAGGACCTGTCTCCACACACTCCCTCTTCCAG 1793
Qy 581 AlaGlnLeuThrAspAspSerAspIleTyrThrAlaGluValAsnGluGluGlyAsp 600
Db 1794 GCCAGCTCAGATGACTCAGACATCTACTGGCCGACAGGTCAACGAGGAGGTGAC 1853
Qy 601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
Db 1854 ACAGTGTCTGTCCCTGAAGAGTTCCTGAAGCAGGATACATATGACGTGCACCTTCT 1913
Qy 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
Db 1914 CTGTCTGACCATGGCAACAAGAGCAGCTACCGTGTATCAGGCGCACTGTGTGCGACTGC 1973
Qy 641 HisGlyHisValGluThrCysProGlyProThrLysGlyGlyPheIleLeuProValLeu 660
Db 1974 CATGGCCATGTGGAACCTGCCCTGGACCCCTGGAAAGGAGTTTCATCTCCCTGTGCTG 2033
Qy 661 GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuValLeuValArgLysLys 680
Db 2034 GGGCTGTCTGCTGTCTGTCTCTCTCTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 2093
Qy 681 ArgLysIleLysGluProLeuLeuProGluAspThrArgAspAsnValPheTyr 700
Db 2094 CGGAAGATCAGGAGGCCCTCTCTACTCCAGAGATGACACCCGTGACACGCTTCTTCTAC 2153
Qy 701 TyrGlyGluGlyGlyGlyGlyGluGluAspLysPheThrGlnLeuHisArg 720
Db 2154 TATGGCAAGAGGGGGTGGCAAGAGGACGAGCTATGACATCACCGAGTCCACCGA 2213
Qy 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
Db 2214 GGTCTGAGGCCAGCGCGAGGTGGTTCTCCGAATGACGTGGCACCACCATCTCCCG 2273
Qy 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleLeuGlu 760

Db 2274 ACACCCATGTCCTGCTAGGCCAGCCACCAACCCAGATGAAATCGGCACTTTATAATTGAG 2333
Qy 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
Db 2334 AACCTGAAGGGCGCTAAACACACAGACCCACGCGCCCTACGACACCCCTCTGTGTTC 2393
Qy 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
Db 2394 GACTATGAGGGGAGGGGTCCGACCGCGCTCCCTGAGCTCCCTCACCTCTCCGCCCTCC 2453
Qy 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTyrGlySerArgPheLysLysLeuAla 820
Db 2454 GACCAAGACCAAGATTACGATTATCTGAACGAGTGGGCGACGCGCTTCAAGAGCTGCCA 2513
Qy 821 AspMetTyrGlyGlyGlyGlyGluAspAsp 829
Db 2514 GACATGTACGTGGGGGAGGAGCAGAC 2540
RESULT 7
ID ABZ24736 standard; cDNA; 3171 BP.
AC ABZ24736;
XX 07-APR-2003 (first entry)
XX Human P-cadherin cDNA.
XX P-cadherin; human; colon cancer; colorectal cancer; cytostatic;
KW Gene therapy; gene; ss.
XX Homo sapiens.
XX Key Location/Qualifiers
FT CDS 54..2543
FT /*tag= a
FT /product= "Human P-cadherin"
XX
PN W0200297395-A2.
XX
PD 05-DEC-2002.
XX
PF 31-MAY-2002; 2002WO-US017109.
XX
PR 31-MAY-2001; 2001US-0294225P.
XX
XX (CHIR) CHIRON CORP.
XX
XX Reinhard C, Klinger J, Jefferson AB, Escobedo J, Randazo F;
PI Winter J, Goodson R;
XX
DR WPI; 2003-140501/13.
DR P-PSDB; ABP58357.
XX
PT Inhibiting migration and proliferation of P-cadherin expressing cancer
PT for treating cancer, especially digestive cancer, characterized by
PT overexpression of P-cadherin, involves administering a P-cadherin
PT antagonist.
XX
PS Disclosure; Page 11-13; 129pp; English.
XX
XX The present sequence is the nucleic acid sequence for human P-cadherin
CC (placental cadherin), a calcium-dependent cellular adhesion protein. The
CC invention provides methods of treating or diagnosing cancers involving P-
CC cadherin expression using ligands that target P-cadherin, especially
CC human anti-P-cadherin antibodies. A claimed method of treating a cancer
CC characterised by the overexpression and/or upregulation of P-cadherin
CC comprises the administration of a P-cadherin antagonist, optionally
CC conjugated to a therapeutic agent. The migration, adhesion and/or
CC proliferation of the cancer is inhibited, and the method is especially
CC useful for treating or preventing a digestive cancer such as colon or
CC colorectal cancer. The antagonists may be an anti-P-cadherin antibody or
CC its fragment, a ribozyme or antisense oligonucleotide. A transgenic

Db 131 TCGCGCGCCTCCGAGCGGTGCGCGCGGTCTTCAGGGAGGCTGAAGTGACCTTGGAGGCG 190
Qy 41 GlyGlyAlaGluGlnProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
Db 191 GGAGCGCGGAGCAGAGCGCGCGCGCGCTGGGAAAGTAATTCATGGCTGGCCTGGG 250
Qy 61 GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValAlaAsnGlyGluThr 80
Db 251 CAAGAGCGAGCTCTGTTAGCACTGATAATGATGACTTCACTGTGGCGAATGGCGAGACA 310
Qy 81 ValGlnGluArgArgSerLeuLysGluArgAsnProLeuLysIlePheProSerIleArg 100
Db 311 GTCCAGGAAGAAGGTCTACTGAAGAAAGGAATCCATTGAAGATCTTCCCATCCAAAGCT 370
Qy 101 IleLeuArgArgHisLysArgAspTrpValValAlaProIleSerValProGluAsnGly 120
Db 371 ATCTTACGAAGACACACAGAGAGATTGGGTGGTTCCTCCAAATATCTGTCCTGAAAATGGC 430
Qy 121 LysGlyProPheProGluArgLeuAsnGlnLeuLysSerAsnLysAspArgAsnThrLys 140
Db 431 AAGGTCCTCTCCCGCCAGAGACTGAATCAGCTCAAGTCTAATAAAGATAGAGACCAAG 490
Qy 141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProGluGlyValPheAlaVal 160
Db 491 ATTTTCTACAGCATCACGGGCGCGGGGCGACAGCCCCCTGAGGGTGTCTTCGCTGTA 550
Qy 161 GluLysGluThrGlyTrpLeuLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys 180
Db 551 GAGAAGGAGACAGAGCTGGTGTGTTGAATAAGCCACTGGACCGGAGGAGATTGCCAAG 610
Qy 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
Db 611 TATGAGCTCTTTGGCCAGCTGTGTCAAGAAATGGTGCCTCAGTGGAGGACCCCATGAAC 670
Qy 201 IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
Db 671 ATCTCCATCATCGTGACCGCACAGATGACACACAGCCCCAAGTTTACCCAGGACACCTTC 730
Qy 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
Db 731 CGAGGGAGTGTCTTAGAGGGAGTCTTACCAGTACTTCTGTGATGACAGGTGCACGCCACG 790
Qy 241 AspGluAspAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln 260
Db 791 GATGAGGATGATGCATCTACACCTACATGGGGTGGTGTGTTACTCATCCATAGCCAA 850
Qy 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
Db 851 GAACCAAGGACCCACACAGCTCATGTTCACCATTCACCGAGCACAGGCACCATCAGC 910
Qy 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
Db 911 GTCATCTCCAGTGGCCTGGACCGGGAAGATCCCTCAGTACACACTGACATCCAGGCC 970
Qy 301 ThrAspMetAspGlyAspGlySerThrThrAlaValAlaValValGluIleLeuAsp 320
Db 971 ACAGACATGATGGGACGGCTCCACACACCGACGTGGCAGTGTAGTGAGATCCTTGAT 1030
Qy 321 AlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
Db 1031 GCCAATGACAAATGCTCCATGTTTGAACCCCGCAGAAAGTACAGAGCCCATGTGCGCTGAGAT 1090
Qy 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
Db 1091 GCAGTGGGCCATGAGGTGCAGAGGCTGACGGTCACTGATCTGGACGCCGCCCACTCACCA 1150
Qy 361 AlaTrpArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
Db 1151 GCCTGGCGGTGCCACTTATCATGGCGGTGACGACGGGACCATTTTACCATCACC 1210
Qy 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
Db 1211 ACCACCCCTGAGAGCAACACAGGGCATCTCTGACAAACAGGAAGGTTTGGATTTTGGAGGC 1270

Qy 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db 1271 AAAAACCCAGACACCCCTGTACGTGAAGTGACCAACAGAGCCCTTTTGTGCTGAAGCTC 1330
Qy 421 ProThrSerThrAlaThrIleValHisValGluAspValAsnGluAlaProValPhe 440
Db 1331 CCACCTCCACACACCCACCATAGTGGTCCAGCTGGAGAGTGAATGAGGACCTGTGTTT 1390
Qy 441 ValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
Db 1391 GTCCACCCCTCCAAAGTCGTTGAGGTCCAGGAGGCGATCCCCACTGGGAGCCGTGTGT 1450
Qy 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db 1451 GTCTACACTGCAGAAAGCCCTGCAAGAGAGATCAAAAGATCAGCTACCCTCCCTGAGA 1510
Qy 481 AspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db 1511 GACCCAGCAGGGTGGCTAGCCATGGACCCAGACAGTGGCGAGGTACACAGTGTGGGCACC 1570
Qy 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db 1571 CTCGACCGTGAGGATGAGCAGTTTGTGAGGAACAACATCTATGAAGTCAATGCTTGGGC 1630
Qy 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAsp 540
Db 1631 ATGGACATGGAAGCCCTCCACCACTGGACCGGGAACCTTCTGCTAACACTGATTGAT 1690
Qy 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 1691 GTCAATGACCATGGCCCGCTGAGCCCGCTCAGATCACCATCTGCAACCAAGCCCTTCCAG 1750
Qy 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
Db 1751 GTGGCGCAGGTGCTGACATCAGGACAAAGACCTGTCTCCCAACACCTCCCTTCCAG 1810
Qy 581 AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGlyAsp 600
Db 1811 GCCAGCTCACAGATGACTCAGACATCTACTGGACGGCAGAGGTCAACAGGAAAGGTGAC 1870
Qy 601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
Db 1871 ACAGTGGTCTTGTCCCTGAAGAAGTCTCTGAAGCAGGATACATATGAGGTGACCTTCT 1930
Qy 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
Db 1931 CTGTCTGACCATGGCAACAAAGAGCAGCTGACGGTGTATCAGGGCCACTGTGTGCGACTGC 1990
Qy 641 HisGlyHisValGluThrCysProGlyProThrLysGlyGlyPheIleLeuProValLeu 660
Db 1991 CATGGCCATGTCCAAACCTGCCCTGGACCTCGAAGGAGGTTTCACTCCCTCTGTGTG 2050
Qy 661 GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysLys 680
Db 2051 GGGGCTGTCTGGCTCTGCTGTCTCTCTGCTGGTGTGCTTGTGTTGTTGGTGAAGAAGAG 2110
Qy 681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
Db 2111 CGGAAGATCAGAGAGCCCTCTCTACTCCAGAAAGATGACACCCGTGACAACTCTCTTAC 2170
Qy 701 TyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
Db 2171 TATGGCGAAGAGGGGGTGGCGAAGGACCAAGGACTATGACATCACCAGCTCCACCGA 2230
Qy 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
Db 2231 GGTCTGGAGGCCAGCGGAGGTGTTCTCCGCAATGACGTGGCACCACCAACCATCATCCG 2290
Qy 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
Db 2291 ACACCATGTACCGTCCTCGCGCCAGCAACCCAGATGAATCGGAACCTTTATATATTGAG 2350

761	AsnLeuLysAlaAlaAenThrAspProThrAlaProProTyrAspThrLeuLeuValphe	780
2351	AACTGTAAAGCGGCTAACACAGACCCACAGCGCCCGCTACGACACCTCTTGGTGTTC	2410
781	AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer	800
2411	GACTATAGGGGACGGGCTCCGACGGCGGTCCCTGAGTCCCTCACCTCTCGCGCTCC	2470
801	AspGlnAspGlnAspTyrAspTyrLeuAsnGluTTPGlySerAArgpheLysLysLeuAla	820
2471	GACCAAGACCAAGATTACGATTATCTGACAGTGGGGCAGCCGGTTCAGAAAGCTGGCA	2530
821	AspMetTyrGlyGlyGluAspAsp	829
2531	CACATGTACCGTGGCGGGAGGACGAC	2557

RESULT 11

ACC72821 standard: cDNA: 3205 BP.

AC 72821 :

09-JUL-2003 (first entry)

Human cancer related protein encoding cDNA

Human; cancer; diagnosis; screening; modulator; leukaemia; ischaemia; heart disease; atherosclerosis; endometriosis; gene; ss.

XX
XY
SC
Homo sapiens

XX
DW
W0300303E130-83XX
27-MAR-2003

17 SEP 2003 2003W0115029560

17 SEP-2001 2001US-0323469D

20-SEP-2001; 2001US-032388/P.

PR 08-FEB-2002; 2002US-0355145P.

PR 12-APR-2002; 2002US-0372246P.

PA (EOSB-) EOS BIOTECHNOLOGY INC

PI Afar D, Aziz N, Gish KC, He

XX

DR P-PSDB; ABR58670.

PT New genes that are

therapeutic targets for screening drugs for treating these diseases.

XX

comprises the sequence of any of the genes that are up-regulated or down-regulated in specific cancers (e.g. about 1031 genes up-regulated in acute lymphocytic leukemia). ACC72641 to ACC72860 represent cancer related gene nucleotide sequences which encode the proteins given in ABR59521 to ABR58709. Also described: (1) determining the presence or absence of a pathological cell in a patient; (2) an expression vector comprising a nucleic acid molecule described above; (3) a host cell comprising the vector; (4) an isolated polypeptide, which is encoded by the nucleic acid; (5) an antibody that specifically binds the polypeptide of (4); (6) specifically targeting a compound to a pathological cell in a patient by administering to the patient the antibody above; and (7) a drug screening assay. The nucleic acid is useful as diagnostic markers or therapeutic targets. In particular, the nucleic acid is useful for diagnosing a pathology, e.g. cancer, of the bone marrow,

281 ValIleSerSerGlyLeuAspArgGluLysValProGluThrThrLeuThrIleGlnAla 300
 Db GTCTATCTCCAGTGGCTGGACCGGAAAGATCCCTGAGTACACACTGACCATCCAGGCC 970
 301 ThrAspMetAspGlyAspGlySerThrThrThrAlaValAlaValGluIleLeuAsp 320
 Db ACAGACATGGATGGGACGGCTCCACCACCGCAGTGGCAGTAGTGGAGATCCTTGAT 1030
 321 AlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
 Db GCCAATGACATGCTCCCATGTTTACCCCGAGAGTACGAGGCCCATGTGCCGTGAGAT 1090
 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
 Db GCAGTGGCCCATGAGGTGACAGGCTCAGCGTCACTGATCTGGACGCCCCCACTCACCA 1150
 361 AlaTrpArgAlaThrTyrLeuIleMetGlyValAspAspGlyAspHisPheThrIleThr 380
 Db GCGTGGCGTGCCACTACCTTATCATGGCGGTGACGACGGGACCATTTTACCATCACC 1210
 381 ThrHisProGluSerAsnGlnGlyLeuThrThrArgLysGlyLeuAspPheGluAla 400
 Db ACCACCCCTGAGACCAACCGGCATCTGACCAACACGAGGCCCTTTGTGCTGAAGCTC 1270
 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
 Db 1271 AAAAACCCAGACACCCCTGCTGTTGAAGTGAACACGAGGCCCTTTGTGCTGAAGCTC 1330
 421 ProThrSerThrAlaThrIleValValHisValGluAspValAlaAsnGluAlaProValPhe 440
 Db CCRACTCCACAGCCACCATAGTGGTCCAGCTGGAGGATGTGAATGAGGCACCTGTGTTT 1390
 441 ValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
 Db GTCCCACTCCCAAGTCGTGAGTCCAGAGGGCATCCCACTGGGGAGCCCTGTGTGT 1450
 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
 Db GTCTTACACTGCAGAAAGACCTGACAGAGAGATCAAAAGATCAGTACCGCATCTTGAGA 1510
 481 AspProIlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
 Db GAOCACAGCGGTGGCTAGCATGATGACCATGGACAGAGTGGGAGGTCACTGTGGGCACC 1570
 501 LeuAspArgGluAspGluGlnPheValArgAsnIleTyrGluValMetValLeuAla 520
 Db CTGACCGTGAAGTGAAGTGTGTGAGGAACAACATCTATGAAGTCACTGTGCTTGGCC 1630
 521 MetAspAsnGlySerProProThrThrGlyThrLeuLeuLeuThrLeuIleAsp 540
 Db ATGGACAAATGGAAGCCCTCCACCATGGCACGGGAACCTTCTGTAACTGATTTGAT 1690
 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
 Db GTCAATGACCATGGCCAGTCCCTGAGCCCCGTGACATCACTGCAACCAAGACCCCT 1750
 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
 Db GTGGCCAGGTGCTGAACATCAAGGACAGGACCTGTCTCCCACTCCCTTTCAG 1810
 581 AlaGlnLeuThrAspAspSerAspIleTyrThrThrAlaGluValAsnGluGlyAsp 600
 Db GCCAGCTCAAGATGACTCAGACATCTACTGGACCGCAGAGGTCAACGAGGAAGTGAC 1870
 601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
 Db ACAGTGGTCTGTCCCTGAGAGAGTTCCTGAACAGAGATACATATGACGTGACCTTCT 1930
 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
 Db CTGTCTGACCATGGCAACAAAGAGCAGCTGACGGTATCAGGCGCCACTGTGTGCGACTGC 1990

641 HisGlyHisValGluThrCysProGlyProTrpLysGlyGlyPheIleLeuProValLeu 660
 Db 1991 CATGGCCATGTGAAACCTGCTGACCTCGAAGGAGGTTTCATCTCTCCCTGTGCTG 2050
 661 GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuValLeuValArgLysLys 680
 Db 2051 GGGGCTGTCTCGCTCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2110
 681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
 Db 2111 CGGAAGATCAAGAGGCCCTCTCTACTCCAGAAAGATGACACCCGTGACAACTCTTCTAC 2170
 701 TyrGlyGluGluGlyGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
 Db 2171 TATGGCAAGAGGGGGGTGGAGAGGACCAAGGACTATGACATCACCAGCTCCACCGA 2230
 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
 Db 2231 GGTCTGGAGGCCAGGCCGAGGTGGTCTCCGCATATGACGTGGCACCAACCATCATCCG 2290
 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
 Db 2291 ACACCCATGTACCTGCTCTCGGCAGCAACCCAGATGAAATCGCAACTTTATAATTGAG 2350
 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuValPhe 780
 Db 2351 AACCTGAAGCGGCTTAAACACAGACCCACAGCCCCGCCCTACGACACCTCTTGTGTTC 2410
 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
 Db 2411 GACTATGAGGGCAGCGGCTCCGACCGCGTCCCTGAGTCCCTCCTCCCTCCGCTCC 2470
 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGlySerArgPheLysLysLeuAla 820
 Db 2471 GACCAAGACCAAGATTACGATTTATCTGAACGAGTGGGCGAGCCGCTTCAAGAAGCTGGCA 2530
 821 AspMetTyrGlyGlyGlyLysAsp 829
 Db 2531 GACATGTACGTGGGGGGGAGGAGCGAC 2557
 RESULT 12
 ABX76155
 ID ABX76155 standard; DNA; 3205 BP.
 XX
 AC ABX76155;
 XX
 DT 02-APR-2003 (first entry)
 XX
 DE Lung cancer-associated polynucleotide #27.
 XX
 KW Lung cancer-associated polynucleotide; gene; ds; cytostatic; emphysema;
 KW antiinflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;
 KW small cell lung cancer; benign lesion; precancerous lesion; bronchitis;
 KW chronic obstructive pulmonary disease; hypersensitivity pneumonitis;
 KW interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.
 OS Unidentified.
 XX
 PN WO200286443-A2.
 XX
 PD 31-OCT-2002.
 XX
 PF 18-APR-2002; 2002WO-US012476.
 XX
 PR 18-APR-2001; 2001US-0284770P.
 PR 10-MAY-2001; 2001US-0290492P.
 PR 09-NOV-2001; 2001US-0339245P.
 PR 13-NOV-2001; 2001US-0350666P.
 PR 29-NOV-2001; 2001US-0334370P.
 PR 12-APR-2002; 2002US-0372246P.
 XX
 PA (EOSB-) EOS BIOTECHNOLOGY INC.

PI	Aziz N, Murray R;	Db	491	ATTTCTACAGCATCAGGGGCGGGGCGACAGCCCCCTGAGGGTGTCTTCGCTGTA	550
XX					
DR	WPI; 2003-093161/08.	Qy	161	GlulysGluThrGlyTrpLeuLeuLeuAsnLysProLeuAspArgGluGluLeuAlaLys	180
DR	P-PSDB; ABU56434.				
XX		Db	551	GAGAAGGAGACAGAGCTGGTTCTTGAATAAGCCACTGGACCGGGAGGAGATTGCCAAG	610
PT	Detecting a lung cancer-associated transcript in a cell from a patient	Qy	181	TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn	200
PT	for treating lung cancer, by contacting a biological sample from the				
PT	patient with a polynucleotide that exhibits increased or decreased	Db	611	TATGAGCTCTTTGGCCACGCTGTGTACAGAAATGGTGTCTCAGTGGAGGACCCCATGAAC	670
XX	expression in lung cancer.				
PS	Claim 22; Page 209; 453pp; English.	Qy	201	IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe	220
XX	The invention relates to a method for detecting a lung cancer-associated	Db	671	ATCTCCATCATCTGTCAGCCAGCATGACCAAGCCCAAGTTTACCCAGGACACCTTC	730
CC	transcript in a cell from a patient, comprising contacting a biological				
CC	sample from the patient with a polynucleotide that selectively hybridises	Qy	221	ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr	240
CC	to a sequence that is at least 80 % identical to a gene that exhibits	Db	731	CGAGGAGTGTCTTAGAGGGAGTCTTACAGGTACTTCTGTGATGCGAGGTGACAGCCACG	790
CC	increased or decreased expression in lung cancer samples. Lung cancer-				
CC	associated polynucleotides and polypeptides are used for identifying a	Qy	241	AspGluAspAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln	260
CC	compound that modulates a lung cancer-associated polypeptide, for	Db	791	GATGAGGATGATGCCATCTACACCTACATGGGTGGTGGTCTTACTCCATCCATAGCAA	850
CC	inhibiting proliferation of a lung cancer-associated cell to treat lung				
CC	cancer in a patient and for treating a mammal having lung cancer by	Qy	261	GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer	280
CC	administering a modulatory compound identified. The methods are useful	Db	851	GAACCAAGGACCCACACGACCTCATGTCTCACCATTACCGGAGCAGAGGCACCATCAGC	910
CC	for treating lung cancer, such as small cell lung cancer, non-small cell	Qy	281	ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla	300
CC	lung cancer or other benign or precancerous lesions, e.g. atelectasis,	Db	911	GTCACTCTCAGTGGCTCGACCGGAAAAAGTCCCTGAGTACACACTGACCATCAGGCC	970
CC	emphysema, bronchitis, chronic obstructive pulmonary disease, fibrosis,	Qy	301	ThrAspMetAspGlyAspGlySerThrThrAlaValAlaValAlaValGluIleuAsp	320
CC	hyperensitivity pneumonitis, interstitial pulmonary fibrosis, asthma and	Db	971	ACAGACATGGATGGGAGCGGCTCCACCAACCGGAGTGGCAGTAGTGAGATCCTTGAT	1030
CC	bronchiectasis. The genes, polynucleotides and polypeptides are useful				
CC	for diagnostic purposes and as targets for screening for therapeutic	Qy	321	AlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn	340
CC	compounds that modulate lung cancer, such as antibodies. Sequences	Db	1031	GCCAATGACATGCTCCCATGTTTGACCCCGAGAGTACGAGGCCCATGTGCTTGAGAT	1090
CC	ABX76124-ABX76474 represent lung cancer-associated polynucleotides of the	Qy	341	AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro	360
CC	invention	Db	1091	CGAGTGGGCCATGAGGTGCAGAGGCTGACGGTCACTGATCTGGAGCGCCCACTACCA	1150
XX					
SQ	Sequence 3205 BP; 751 A; 909 C; 875 G; 670 T; 0 U; 0 Other;	Qy	361	AlaTrpArgAlaThrTyrIleuMetGlyGlyAspAspGlyAspHisPheThrIleThr	380
		Db	1151	GGTGGCGTGCACCTACCTTATCATGGCGGTGACGACGGGACCATTTTACCATCACC	1210
		Qy	381	ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla	400
		Db	1211	ACCCACCTTGAGAGCAACCCAGGCGCATCTGACACACAGGAAGGGTTTGGATTTTGAGGCC	1270
		Qy	401	LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValIleuLysLeu	420
		Db	1271	AAAAACCCAGCACACCTGTACGTTGAAGTGACCAACAGAGGCCCTTTTGTGCTGAAGCTC	1330
		Qy	421	ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe	440
		Db	1331	CCAACTCCACAGCCACCATAGTGTCCCGTGGAGGATGTGAATGAGGCACCTGTGTTT	1390
		Qy	441	ValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys	460
		Db	1391	GTCCCACTCCAAAAGTCTGTGAGGTCCAGGAGGCGCATCCCCACCTGGGGAGCGCTGTGT	1450
		Qy	461	ValThrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg	480
		Db	1451	GTCTACACTGCAAGACACCTGCAAGGAGATCAAAAGATCAGCTACCGCATCTCGAGA	1510
		Qy	481	AspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr	500
		Db	1511	GACCCAGCAGGTGGCTAGCCATGACCCAGAGTGGGCGAGGTCCACAGCTGTGGGCACC	1570
		Qy	501	LeuAspArgGluAspGluGlnPheValArgAsnIleTyrGluValMetValLeuAla	520

Qy	21	CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla	40	Qy	361	ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgIysGlyLeuAspPheGluAla	400
Db	131	TGCCGCGCTCCGAGCGCTGCGCGCGGTCTTCAGGAGGCTGAGTGACCTTGGAGGCG	190	Db	1211	ACCCACCTTGAGAGCAACAGGGCATCTGACAAACAGGAGGGTTTTGGATTTTGGAGGCC	1270
Qy	41	GlyGlyAlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly	60	Qy	401	LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu	420
Db	191	GGAGCGCGGAGCAGGAGCGCGCGCGCTGGGGAAGTATTCATGGGCTGCCCTGGG	250	Db	1271	AAAAACCAAGACACCTGTGTACGTTGAAGTGACCAACGAGGCCCCCTTTTGTGCTGAACCTC	1330
Qy	61	GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr	80	Qy	421	ProThrSerThrAlaThrIleValHisValGluAspValAsnGluAlaProValPhe	440
Db	251	CAAGAGCCAGCTCTGTTTAGCAGTGAATGATGACTTCACTGTGCGGAATGGCAGACA	310	Db	1331	CCACCTCCACAGCCACCATAGTGTCCACGTGAGGATGTGAATGAGGACCTGTGT	1390
Qy	81	ValGlnGluArgArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg	100	Qy	441	ValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys	460
Db	311	GTCCAGGAGGAAGGTCACTGAAGGAAGGAATCCATTGAAGATCTTCCCATCAACCGT	370	Db	1391	GTCCACACCTCCAAAGTCGTTGAGGTCAGGAGGCAATCCCACTGGGAGCCTGTGT	1450
Qy	101	IleLeuArgArgHisLysArgAspTyrValValAlaProIleSerValProGluAsnGly	120	Qy	461	ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg	480
Db	371	ATCTTACGAGACACAAGAGATTGGGTGGTTGCTCCAAATATCTGTCCCTGAAATGGC	430	Db	1451	GTCTACACTGCGAAGACCTTGACAGGAGAATCAAAAGATCAGCTACCGCATCTCGAGA	1510
Qy	121	LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys	140	Qy	481	AspProAlaGlyTyrPheAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr	500
Db	431	AAGGTCCTCTTCCCCAGAGACTGAATCAGCTCAAGTCTAATAAGATAGAGACCAAG	490	Db	1511	GACCCAGCAGGCTGGCTAGCCATGACCCACAGCAGTGGGCGAGGTCACAGCTGTGGGACC	1570
Qy	141	IlePheTyrSerIleThrGlyProGlyValAlaAspSerProGluGlyValPheAlaVal	160	Qy	501	LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla	520
Db	491	ATTTTCTACAGCATCACGGGCGCGGGGCGAGACAGCCCTCTGAGGGTGTCTTGGCTGTA	550	Db	1571	CTCGACCGTGAGGATGAGCAGTTTGTGAGGAACAACATCTATGAAGTCATGGTCTTGCC	1630
Qy	161	GluLysGluThrGlyTyrPheLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys	180	Qy	521	MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuLeasp	540
Db	551	GAGAGCAGAGCAGGCTGGTGTGTGTAATGAAGCCATGGACCGGAGAGATGCCAG	610	Db	1631	ATGGCAATGGAAGCCCTCCACCACTGGCAGCGGGAACCTTCTGCTAACACTGATTGAT	1690
Qy	181	TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn	200	Qy	541	ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro	560
Db	611	TATGAGCTCTTTGGCCACGCTGTGTGAGAGAATGGTGCCTCAGTGGAGGACCCCATGAAC	670	Db	1691	GTCAATGACATGGCCCAAGTCCCTGAGCCCGGTGAGATCACCACTGCAACCAAGCCCT	1750
Qy	201	IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe	220	Qy	561	ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln	580
Db	671	ATCTCCATCATCGTACCGACAGAAATGACCACAAGCCCAAGTTTACCAGGACACCTTC	730	Db	1751	GTGGCCAGGTGCTGAACATCAGCGACAAGGACCTGTCTCCCCACACCTCCCTTTCCAG	1810
Qy	221	ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr	240	Qy	581	AlaGlnLeuThrAspAspSerAspIleTyrTyrThrAlaGluValAsnGluGluGlyAsp	600
Db	731	CGAGGAGTGTCTTAGAGGAGTCTTACAGAGTACTTGTGTGATGACAGGTGACGCCACG	790	Db	1811	GCCAGCTCACAGATGACTCAGACATCTACTGACCGCAGAGGTCAACAGGAAGGTGAC	1870
Qy	241	AspGluAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln	260	Qy	601	ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer	620
Db	791	GATGAGGATGATGCCATACACCTACACCTACCAATGGGTGGTGTCTTACTCCATCCATAGCAA	850	Db	1871	ACAGTGGTCTTGTCCCTGAAGAAGTTCCTGAAGCAGGATACATATGACGTGACCTTTCT	1930
Qy	261	GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer	280	Qy	621	LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys	640
Db	851	GAACCAAGGACCCACAGACTCATGTTTACCATTACCCGAGCAGCAGGACCATCAGC	910	Db	1931	CTGTCTGACCATGGCAACAAGAGCAGCTGACGGTGATCAGGGCCACTGTGTGCCACTGC	1990
Qy	281	ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGluAla	300	Qy	641	HisGlyHisValGluThrCysProGlyProTyrLysGlyGlyPheIleLeuProValLeu	660
Db	911	GTCACTCCAGTGGCTTGACCGGGAAGATCCCTGAGTACACACTGACCATCCAGGCC	970	Db	1991	CATGGCCATGTGAAACCTGCCCTGGACCCCTGAGAGGAGGTTTCATCTCCTCTGTGTG	2050
Qy	301	ThrAspMetAspGlyAspGlySerThrThrThrAlaValAlaValGluIleLeuAsp	320	Qy	661	GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysLys	680
Db	971	ACAGACATGATGGGAGCGCTCCACCAACCGCAGGTGGCAGGTAGTGAGATCTTGAT	1030	Db	2051	GGGGCTGTCTGGCTGTCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	2110
Qy	321	AlaAsnAspAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn	340	Qy	681	ArgLysValIleGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr	700
Db	1031	GCCATGACAAATGCTCCCATGTTTACCCCCCAGAAAGTACGAGGCCCATGTGCTGAGAAT	1090	Db	2111	CGGAAGATCAGAGAGCCCTCTCTCTCCCAAGAGATGACACCCGTCGACACGCTCTTCTAC	2170
Qy	341	AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro	360	Qy	701	TyrGlyGluGluGlyGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisArg	720
Db	1091	GCAGTGGGCGCATGAGGTGAGAGGTGACGGTCACTGATCTGGACGCCCCCACTCACC	1150	Db	2171	TATGGCAAGAGGGGGTGGCAAGAGGACAGGACTATGACATCACCCAGTCCACCGA	2230
Qy	361	AlaTyrArgAlaThrTyrIleIleMetGlyGlyAspAspGlyAspHisPheThrIleThr	380	Qy	721	GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro	740
Db	1151	GGTGGCGTGGCCACCTACCTTATCATGGCGGTGACGAGGGACCACTTTTACCATCACC	1210	Db	2231	GGTCTGGAGCCAGGCGGAGGTGTTTCTCCGCAATGAGCTGGCACCACCACTATCCCG	2290
Qy				Qy	741	ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu	760

```
Db 2291 ACACCCATGTACCTCTCGGCCAGCCACCATGAAATCGCACTTTATAATTGAG 2350
Qy 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProThrAspThrLeuLeuValPhe 780
Db 2351 AACCTGAAGCGGCTAACACAGACCCACAGCCCGCCCTACGACACCCCTCTGGTGTC 2410
Qy 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerAlaSer 800
Db 2411 GACTATAGGCGACGCGCTCCGACGCGCGTCCCTGAGCTCCCTCACCTCTCCGCTCC 2470
Qy 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGlySerArgPheLysLysLeuAla 820
Db 2471 GACCAAGACCCAGATTACGATTATCTGAACGAGTGGGGCAGCGCTTCAAGAGCTGGCA 2530
Qy 821 AspMetTyrGlyGlyGlyGluAspAsp 829
Db 2531 GACATGTACGCTGGCGGAGGACGAC 2557
RESULT 14
AAD27637
ID AAD27637 standard; DNA; 3219 BP.
XX AC
AC AAD27637;
XX
XX 18-APR-2002 (first entry)
XX Human colon specific gene #17.
XX Human; colon specific gene; CSG; imaging; colon cancer; metastasis;
XX screening; chromosomal identification; transgenic animal; cytostatic;
XX gene therapy; p cadherin; ds.
XX Homo sapiens.
XX W0200200939-A2.
XX
XX 03-JAN-2002.
XX 28-JUN-2001; 2001WO-US020724.
XX
XX 28-JUN-2000; 2000US-0214515P.
XX (DIAD-) DIADEXUS INC.
XX Macina RA, Pillai R;
XX WPI; 2002-139938/18.
XX
XX Detecting, diagnosing, monitoring, staging, prognosticating, imaging and
XX treating colon cancer using colon-specific genes.
XX Claim 1; Page 127-128; 135pp; English.
XX
XX The invention relates to colon specific gene (CSG) and their polypeptides
XX used in a method of diagnosing, monitoring, staging, imaging and treating
XX colon cancer. CSG is useful for diagnosing the presence and metastasis of
XX colon cancer. It can be used for staging colon cancer in a patient having
XX colon cancer where an increase in determined CSG levels in the patient
XX compared to normal human control is associated with a cancer which is
XX progressing and a decrease in CSG levels is associated with a cancer
XX which is regressing or in remission. CSG is useful for identifying
XX potential therapeutic agents for use in imaging and treating colon cancer
XX by screening compounds that binds to, or decrease expression of CSG
XX relative CSG in the absence of the compound. CSG is used to identify
XX subjects having or at risk of developing a disease or disorder associated
XX with increased levels of CSG. CSG sequences are also used for chromosomal
XX identification. CSG sequences can also be used for producing non-human
XX transgenic animals. The methods can also be used to detect genetic
XX lesions or mutations in CSG. The identification of CSG is also useful in
XX the rational design of the new therapeutics for imaging and treating
XX cancers. CSG is used in gene therapy. The present sequence is human colon
XX specific gene encoding p cadherin
```

```
XX
SQ Sequence 3219 BP; 748 A; 915 C; 884 G; 672 T; 0 U; 0 Other;
Alignment Scores:
Pred No.: 0 Length: 3219
Score: 4348.00 Matches: 828
Percent Similarity: 99.76% Conservatives: 0
Best Local Similarity: 99.76% Mismatches: 1
Query Match: 99.52% Indels: 1
DB: 6 Gaps: 0
US-09-916-849a-1 (1-829) x AAD27637 (1-3219)
Qy 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuLeuValCysTrpLeuGln 20
Db 90 ATGGGGCTCCCTCGTGGACCTCTCGCTCTCTCCCTCTCTCCAGGTTTGTGGGTGCGAC 149
Qy 21 CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
Db 150 TCGGGGGCTCCGAGCCGTCGCGGGGGTCTTCAGGGAGGCTGAAGTGACCTTGGAGGG 209
Qy 41 GlyAlaGluGluGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
Db 210 GGAGGCGGAGCAGAGAGCCGCGCCAGCGCTGGGGAAGATTATTCATGGGCTGCCCTGG 269
Qy 61 GlnGluProAlaLeuPheSerThrAspAsnAspAspPheThrValArgAsnGlyGluThr 80
Db 270 CAAGAGCCAGCTCTCTTTAGCACTGATATGATGACTTCACTGCGGAATGGCAGACA 329
Qy 81 ValGlnGluArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
Db 330 GTCCAGGAAGAAGGTCCTCTGAAGGAAGAAGTAATCATTTGAAGATCTTCCCATCCAACT 389
Qy 101 IleLeuArgArgHisLysArgAspTrpValAlaAlaProLleSerValProGluAsnGly 120
Db 390 ATCTTACGAAGACACAAGAGAGATTGGGTGGTGTCTCCAATATCTGCTCCCTGAAATGGC 449
Qy 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
Db 450 AAGGTGCTCTTCCCGCAGAGACTGAATCAGCTCAAGTCTAATAAGATAGAGACACCAAG 509
Qy 141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProProGluGlyValPheAlaVal 160
Db 510 ATTTTCTACGATCACGGGGCGGGGCGAGACAGCCCCCTGTAGGGTGTCTTCTCGTGT 569
Qy 161 GluLysGluThrGlyTrpLeuLeuAsnLysProLeuAspArgGluLysIleAlaLys 180
Db 570 GAGAGGAGACAGGCTGTTGTTGTTGAATAGCCACTGGACCGGGAGAGATTGCCAAG 629
Qy 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
Db 630 TATGAGCTCTTTGGCCACGCTGTGTGACAGAAATGGTGCTCAGTGGAGGACCCCATGAAC 689
Qy 201 IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
Db 690 ATCTCCATCATGCTGACCGACAGAAATGACCACAAGCCCAAGTTTACCAGGACACCTTC 749
Qy 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
Db 750 CGAGGGAGTGCTTAGAGGGAGTCTTACAGGACTCTTCTGTGATGTCAGGTGACGCCACA 809
Qy 241 AspGluAspAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln 260
Db 810 GATGAGGATGATGCCATCTACACCTACCAATGGGTGGTGTGCTTACTCCATCCATAGCCAA 869
Qy 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
Db 870 GAACCAAGAGCCACACAGACTCATGTTCAATTCACCGGAGCAGAGGACCATCAGC 929
Qy 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
Db 930 GTCATCTCCAGTGGCTGGACCGGGAAAAAGTCCCTGAGTACACACTGACCATCCAGGCC 989
```


Qy	578	oPheGlnAlaGlnLeuThrAspSerAspIleTyTrpThrAlaGluValAsnGluGl	598
Db	1953	CTTTCACAGCAGAACTATACACACGCGCAAGTGTCAACTGACCATCGAGTACAAATGACCC	2012
Qy	598	uGly---AspThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrraspVa	617
Db	2013	AGTCGTGATCTCTAATTTGAAGCCAAAGAAAACCTTTAGAGTTGGTGACTACAAAAT	2072
Qy	617	lHisLeuSerLeuSerAspHisGlyAsnLysGluGlnLeuThrValIleAargAlaThrVa	637
Db	2073	AAATCTCAAGCTCACAGATACACGAAACAAGGACCAAGTGCACCACTATATGTGTGTGT	2132
Qy	637	lCysAspCysHisGlyHisValGluThrCys-----ProGlyProTrp---LysGlyGl	654
Db	2133	GTGCGACTGGAGGTGCTGTCACAGCTGCAAGAGGACGGCGCTTACGCCGAGCAGG	2192
Qy	654	yPhe-----IleLeuProValLeuGlyAlaValLeuAlaLeuLeuPheIleuLe	670
Db	2193	CTTGCGAGTTCCTGCCACTCTGGGCATCTCGGAGGAATCTCGCTCTACTAATCTCTGAT	2252
Qy	670	uLeuValLeuLeuLeuValArgLysLysAsqLysIleLysGlnProLeuLeuLeuPr	690
Db	2253	TCGTCTGCTTCGTCTATTGTTTCGGAGGAGAGGGTGTCAAGAGCCCTTACTTCCCCC	2312
Qy	690	oGluAspAspThrArgAspAsnValPheTyTyrrGlyGluGluGlyGlyGluGluAs	710
Db	2313	AGAAGATGATCACCGCGGACAAATGTTATTACTATGATGAAGAAGAGGAGTGGAGAGGAGGA	2372
Qy	710	pGlnAspTyrrAspIleThrGlnLeuHisAsqGlyLeuGluAlaAargProGluValValle	730
Db	2373	TCAGGACTTTGACTGTGAGCCAGTTGCACAGGGCCCTGGATGCTCGGCTCGAAGTACT--	2430
Qy	730	uArgAspValAlaProThrIleIleProThrProMetTyrrArgProAargProAlaAs	750
Db	2431	-CGCAATGATGTGGCCCCAACCCCTCTCGAGTGTGGCCAGTATCGGCCCGCCCTGCGCAA	2489
Qy	750	nProAspGluIleGlyAsnPheIleIleGluAsnLeuLysAlaAlaAsnThrAspProTh	770
Db	2490	TCCTGATCAATTTGGAACCTTTATTGATGAACCTGAAGCGAGCGGACACTGACCTTAC	2549
Qy	770	rAlaProProTyrrAspThrLeuLeuValPheAspTyrrGluGlySerGlySerAspAlaAl	790
Db	2550	TGCTCTCTCTTATGACTCTCTGCTCGTGTGTGACTATGAAGGAAGCGGGTTCGGAAGCTGC	2609
Qy	790	aSerLeuSerSerLeuThrSerSeralAspGlnAspGlnAspTyrrAspTyrrIleuAs	810
Db	2610	TAGTCTGAGCTCTCTGAACTCTCTCAGAGTCAGACCAACAGCAGGACTAGACTTACCTGAA	2669
Qy	810	nGluTrpGlySerArgPheLysLysLeuAlaAspMetTyrrGlyGlyGluAspAsp	829
Db	2670	TGAATGGGCAATCGCTTCAAGAAGCTGGCGGCATGTATGGAGGTGGCAGAGCAGCAC	2727

Search completed: September 22, 2004, 21:44:54
Job time : 864 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 22, 2004, 21:23:38 ; Search time 142 Seconds
(without alignments)

3239.819 Million cell updates/sec

Title: US-09-916-849a-1

Perfect score: 4369

Sequence: 1 MGLPRGLASLLLVQCWLQ.....NEWGSRFKLLADMYGGEDD 829

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus_p2n.model -DEV=xlh
-C=/cgm2_1/USPTO_spool_h/US0916849/runat_21092004_145804_7033/app_query.fasta_1.967
-DB=Issued Patents NA -OPMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US0916849.cgm2_1_69 @runat_21092004_145804_7033 -NCPU=6 -ICPU=3
-DEV -MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6
-FGAEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:
1: /cgm2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgm2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgm2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgm2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgm2_6/ptodata/2/ina/PCTUS.COMB.seq:*
6: /cgm2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4369	100.0	3171	4	US-09-016-434-1348
2	2349.5	53.8	2808	1	US-08-237-919-1
3	2349.5	53.8	2808	4	US-08-732-429-1
4	2349.5	53.8	2808	4	US-09-798-267-1
5	2349.5	53.8	2808	5	PCT-US95-05518-1
6	1621.5	37.1	3048	1	US-08-188-228-47
7	1621.5	37.1	3048	1	US-08-332-643-41
8	1621.5	37.1	3048	1	US-08-332-638-47
9	1025.5	23.5	2690	1	US-08-188-228-61
10	1025.5	23.5	2690	1	US-08-332-643-55
11	1025.5	23.5	2690	1	US-08-332-638-61
12	977.5	22.4	2779	1	US-08-474-067-3

13	977.5	22.4	2779	2	US-08-474-068A-3	Sequence 3, Appl
14	977.5	22.4	2779	2	US-08-472-481-3	Sequence 3, Appl
15	976.5	22.4	2880	4	US-09-016-434-1125	Sequence 1125, Ap
16	976	22.3	3959	1	US-08-474-067-1	Sequence 1, Appl
17	976	22.3	3959	2	US-08-474-068A-1	Sequence 1, Appl
18	976	22.3	3959	2	US-08-472-481-1	Sequence 1, Appl
19	968.5	22.2	3712	2	US-08-738-349-3	Sequence 3, Appl
20	968	22.2	4125	4	US-09-620-312D-270	Sequence 270, App
21	966.5	22.1	2625	1	US-08-188-228-57	Sequence 57, Appl
22	966.5	22.1	2625	1	US-08-332-643-51	Sequence 51, Appl
23	966.5	22.1	2625	1	US-08-332-638-57	Sequence 57, Appl
24	964.5	22.1	3552	4	US-09-643-597-126	Sequence 126, App
25	964.5	22.1	3552	4	US-09-480-884A-126	Sequence 126, App
26	964.5	22.1	3552	4	US-09-542-615A-126	Sequence 126, App
27	964.5	22.1	3552	4	US-09-606-421B-126	Sequence 126, App
28	964.5	22.1	3552	4	US-09-221-107-126	Sequence 126, App
29	951.5	21.8	3581	2	US-08-738-349-1	Sequence 1, Appl
30	933	21.4	3136	1	US-08-188-228-41	Sequence 41, Appl
31	933	21.4	3136	1	US-08-332-638-41	Sequence 41, Appl
32	926.5	21.2	2550	1	US-08-188-228-53	Sequence 53, Appl
33	926.5	21.2	2550	1	US-08-332-643-47	Sequence 47, Appl
34	926.5	21.2	2550	1	US-08-332-638-53	Sequence 53, Appl
35	924	21.1	3514	2	US-08-738-349-5	Sequence 5, Appl
36	914.5	20.9	2521	1	US-08-188-228-59	Sequence 59, Appl
37	914.5	20.9	2521	1	US-08-332-643-53	Sequence 53, Appl
38	914.5	20.9	2521	1	US-08-332-638-59	Sequence 59, Appl
39	853.5	19.5	2490	1	US-08-188-228-45	Sequence 45, Appl
40	853.5	19.5	2490	1	US-08-332-638-45	Sequence 45, Appl
41	796	18.2	4000	4	US-09-016-434-1434	Sequence 1434, Ap
42	796	18.2	4000	4	US-09-023-655-1452	Sequence 1452, Ap
43	794	18.2	3164	1	US-08-188-228-49	Sequence 49, Appl
44	794	18.2	3164	1	US-08-332-643-43	Sequence 43, Appl
45	794	18.2	3164	1	US-08-332-638-49	Sequence 49, Appl

ALIGNMENTS

RESULT 1

US-09-016-434-1348
; Sequence 1348, Application US/09016434
; Patent No. 6500938

GENERAL INFORMATION:

APPLICANT: Janice Au-Young

APPLICANT: Jeffrey J. Seilhamer

TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING

TITLE OF INVENTION: PATHWAY GENE EXPRESSION

NUMBER OF SEQUENCES: 1490

CORRESPONDENCE ADDRESS:

ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

STREET: 3174 PORTER DRIVE

CITY: PALO ALTO

STATE: CALIFORNIA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/016,434

FILING DATE: HEREWITH

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Zeller, Karen J

REGISTRATION NUMBER: 37,071

REFERENCE/DOCKET NUMBER: PA-0002 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (650) 855-0555

TELEFAX: (650) 845-4166

INFORMATION FOR SEQ ID NO: 1348:

SEQUENCE CHARACTERISTICS:

LENGTH: 3171 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: GENBANK

CLONE: G35322

US-09-016-434-1348

Alignment Scores:

Pred. No.: 0 Length: 3171
Score: 4369.00 Matches: 829
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-09-916-849A-1 (1-829) x US-09-016-434-1348 (1-3171)

QY 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuGlnValCysTrpLeuGln 20
DB 54 ATGGGGCTCCCTCGTGGACCTCGCGTCTCTCCCTCTCCAGGTTTGTGGCTGCAG 113
QY 21 CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
DB 114 TCGCGGGCTCCCGAGCGCTCGCGGCTCTTCAGGGAGGCTGAAGTGAACCTTGGAGGCG 173
QY 41 GlyGlyAlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
DB 174 GGAGCGCGGAGCAGAGCGCGCGCGCGCTGGGGAAGTATTTCATGGGCTGCGCTGGG 233
QY 61 GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr 80
DB 234 CAAGAGCCAGCTCTGTTTAGCACTGATTAATGACTTCACTGTCGGAATGGCGAGACA 293
QY 81 ValGlnGluArgArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
DB 294 GTCCAGGAAGAAGGTCACTGAAGAAAGGATTCATTGAAGATCTTCCCATCCAAACGT 353
QY 101 IleLeuArgArgHisLysArgAspTrpValValAlaProIleSerValProGluAsnGly 120
DB 354 ATCTTACGAAGACACAGAGAGATTGGGTGTTCTCCTCAATATCTGTCCTCGAAATGGC 413
QY 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
DB 414 AAGGTCCTTCCCGCAGAGACTGAATCAGCTCAAGTCAAGTCAATTAAGATAGACACCAAG 473
QY 141 IlePheTrpSerIleThrGlyProGlyAlaAspSerProGluGlyValPheAlaVal 160
DB 474 ATTTCTACAGCATCACGGGCGCGGGGCGAGACAGCCCCCTGAGGGTGTCTCGCTGTA 533
QY 161 GluLysGluThrGlyTrpLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys 180
DB 534 GAGAAGGAGACAGGCTGGTGTGTTGAATAAGCCACTGGACCGGGAGGAGATTGCCAAG 593
QY 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
DB 594 TATGAGCTCTTTGGCCACGCTGTGTACAGAAATGGTGCCTCAGTGGAGGACCCCATGAAC 653
QY 201 IleSerIleIleValThrAspGluAsnAspHisLysProLysPheThrGlnAspThrPhe 220
DB 654 ATCTCCATCATCTGACGACGACCAAGATGACCAAGCCCAAGTTTATCCCGAGCACCTTC 713
QY 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
DB 714 CGAGGAGTGTCTTAGAGGAGTCTCTACAGGTACTTCTGTGATGCAAGTGCAGCCACA 773
QY 241 AspGluAspAlaIleTrpThrTyAsnGlyValValAlaTrpSerIleHisSerGln 260
DB 774 GATGAGGATGATGCATCTACACCTACAATGGGGTGTGTTACTTACTCCATCCATAGCCAA 833

QY 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
DB 834 GAACCAAGAGGACCCACACGACCTCATGTTCAATTCACCGAGCACAGGACCATCAGC 893
QY 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyThrLeuThrIleGlnAla 300
DB 894 GTCACTCTCCAGTGGCTGGACCGGAAAGTCCCTGAGTACACACTGACCATCCAGGCC 953
QY 301 ThrAspMetAspGlyAspGlySerThrThrThrAlaValAlaValAlaValIleLeuAsp 320
DB 954 ACAGACATGATGGGACGCTCCACCAACGCGAGTGGCAGTAGTGGAGATCTTCAT 1013
QY 321 AlaAsnAspAlaProMetPheAspProGlnLysTyThrGluAlaHisValProGluAsn 340
DB 1014 GCCAATGACAATGCTCCCATGTTGACCCCCAGAAAGTACGAGGCCCATGTGCTGAGAAT 1073
QY 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
DB 1074 GCAGTGGGCGCATGAGTGCAGAGGCTGACGTCATCTGCGAGCGCCCCCAACTCACA 1133
QY 361 AlaTrpArgAlaThrTyThrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
DB 1134 GCGTGGCGTGCACCTACCTTATCATGGCGGTGACGACGGGACCATTTTACCATCACC 1193
QY 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
DB 1194 ACCACCTTCAGAGCAACCCAGGGCATCTGACACACAGGAGGGTTCGGATTTTGAGGCC 1253
QY 401 LysAsnGlnHisThrLeuTyThrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
DB 1254 AAAAACCCAGCACACCTGTACGTTCAAGTGACCAACGAGGCGCTTTTGTGCTGAAGCTC 1313
QY 421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
DB 1314 CCAACCTCCACAGCCACCATAGTGTCCAGTGGAGGATGGAATGAGGCGACCTGTGTGTT 1373
QY 441 ValProProSerLysValValGluValGlnGluGlyIleProThrGlyGlyProValCys 460
DB 1374 GTCCACCCCTCCAAAGTCTGTTGAGTCCAGGAGGSCATCCCCACTGGGAGGCTGTGTGT 1433
QY 461 ValTyThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyArgIleLeuArg 480
DB 1434 GTCTACCTGCAGAGACCCCTGCAAGAGAGATCAAAAGATCAGTACCGCATCCTGAGA 1493
QY 481 AspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
DB 1494 GACCCAGCAGGCTGTAGCCATGACCCAGACAGTGGGAGGTGACAGCTGTGGGCACC 1553
QY 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyThrGluValMetValLeuAla 520
DB 1554 CTCGACCTGTAGGATGAGCAGTTCGTGAGGAACAACATCTATGAAGTCATGCTGTGGCC 1613
QY 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuLeuLeuLeu 540
DB 1614 ATGGCAATGGAAGCCCTCCACCACTGGCAGGGAACCTTCTGCTAACACTGATTCAT 1673
QY 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
DB 1674 GTCAACGACCATGGCCAGTCCCTGAGCCCGTCCAGATCAGATCAGTCAACCAAGGCT 1733
QY 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
DB 1734 GTGCGCCACGCTGCTGAACATCAGGACAGGACCTGTCTCCCCACACCTCCCTTTCCAG 1793
QY 581 AlaGlnLeuThrAspAspSerAspIleTyThrAlaGluValAlaAsnGluGlyAsp 600
DB 1794 GCCCAGCTCACAGATGACTCAGACATCTACTGAGCGGACAGAGGTCAACGAGGAAGGTGAC 1853
QY 601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrThrAspValHisLeuSer 620
DB 1854 ACAGTGGTCTTGTCTCCCTGAAGAAGTCTCTGAAGCAGGATACATATGAGCGTGCACCTTCT 1913

QY 98 rlysrArgIleLeuArgArgHisLysArgAspTrpValValAlaProIleSerValProGI 118
DB 531 CTCTCCTGGCCTCAGAACAGAGAGAGACTGGGTATTCTCCATCAGCTGCCAG 590
QY 118 uAsnGlyLysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAs 138
DB 591 AAATGAAAGGCCCATTTCTTAAACCTGGTTCAGATCAATCCACAAAGACAAGA 650
QY 138 pThrLysIlePheThrSerIleThrGlyProGlyAlaAspSerProGluGluValPh 158
DB 651 AGGCAAGGTTTCTACAGCATCAGTGGCCAAAGGAGCTGACACCCCTGTGTGTCTT 710
QY 158 eAlaValGluLysGluThrGlyTrpLeuLeuAsnLysProLeuAspArgGluGluI 178
DB 711 TATTATTGAAGAGAAACAGAGTGGCTGAAGTGACAGAGCTCTGGATAGAGACGCAT 770
QY 178 eAlaLysTrpGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspPr 198
DB 771 TGCCACATACATCTCTCTCTCAGCTGTGTCTCAACCGGAATGCAGTTGAGGATCC 830
QY 198 oMetAsnIleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAs 218
DB 831 AATGGAGATTTTGATCAGGTACCGATCAGATGACAAAGCCCGAATTCACCCAGGA 890
QY 218 pThrPheArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValTh 238
DB 891 GGTCTTTAAGGGTCTGTCTGTAAGGTGCTCTTCCAGGAACCTCTGTGTGAGGTCTAC 950
QY 238 rAlaThrAspGluAspAlaIleThrThrTyrrAsnGlyValValAlaValSerIleHi 258
DB 951 AGCCACAGCGGACGATGATGTAACCATCAATGCCGCCATCGCTTACACCATCT 1010
QY 258 sSerGlnGluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyTh 278
DB 1011 CAGCCACAGATCCTGAGCTCCCTGACAAAATATGTTCCACCATTAACAGGAACAGGAT 1070
QY 278 rIleSerValIleSerSerGlyLeuAspArgGluLysValProGluThrThrLeuThrI 298
DB 1071 CATCAGTGTGTCCACCTGAGGTGGGCGGAGAGATTTCCCTACGTATACCTCTGGTGGT 1130
QY 298 eGluAlaThrAspMetAspGlyAspGlySerThrThrAlaValAlaValValGluI 318
DB 1131 TCACCTGCTGACCTTCAGGTGAGGGGTAAAGCACACACACACACACTGTGATCAGT 1190
QY 318 eLeuAspAlaAsnAspAlaProMetPheAspProGlnLysTrpGluAlaHisValPr 338
DB 1191 CACTGACACCAACCAATATCTCCGATCTTCAATCCACACACAGGTACAGGTGCC 1250
QY 338 oGluAsnAlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAs 358
DB 1251 TGAAACAGAGCTAACGCTGTAATACCACTGAAAGTGAAGTACTGATGCTGATGCCCCAA 1310
QY 358 nSerProAlaTrpArgAlaThrTrpLeuIleMetGlyLysAspGlyAspHisPheTh 378
DB 1311 TACCCAGCGTGGGAGGCTGATACCATATTG---AATGATGATGGTGGACAATTTG 1367
QY 378 rIleThrThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPh 398
DB 1368 CGTCACCAACAATCAGTGAAACAGATGGATTTTGAACAAACAGCAAGGGCTGGATTT 1427
QY 398 eGluAlaLysAsnGlnHisThrLeuTrpValGluValThrAsnGluAlaProPheVal 418
DB 1428 TGAGCCCAAGCAGCAGTACATCTTACACGTAGCAGTACCGAATGTGTTGAGGT 1487
QY 418 uLysLeuProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaPr 438
DB 1488 CTCCTCTACCACTCCACAGCCAGCCGTCACCGTGGATGTCTGGATGTGAATGAAGGCC 1547
QY 438 oValPheValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluPr 458
DB 1548 CATCTTTGTGCTCTGAAAGAGAGTGGAGTGTCCGAGGACTTTTGGCGTGGCCAGGA 1607

QY 458 oValCysValTrpThrAlaGluAspProAspLys---GluAsnGlnLysIleSerTyrrAr 477
DB 1608 AATCACATCTTACATGCCCAGAGAGCCAGACACATTTATGGAAAGAAAATACATATCG 1667
QY 477 gIleLeuArgAspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAl 497
DB 1668 GATTTGGAGAGACACTCGCAACTGGCTGGAGATTAAATCGGACACTGGTGCCATTCCAC 1727
QY 497 aValGlyThrLeuAspArgGluAspGluGlnPheValArgAsnAsnIleTrpGluValMe 517
DB 1728 TCGGGCTGAGCTGGACAGGGAGATTTTGAGCAGCTGGAAGAACAGCAGCTACACAGCCCT 1787
QY 517 tValLeuAlaMetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuTh 537
DB 1788 AATCATAGCTACAGACAATGGTTCTCCAGTTCTACTGGAACAGGACACTTCTCTGAT 1847
QY 537 rIleuLeAspValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAs 557
DB 1848 CTTGTCTGATGTGAATGACACGCCGCCCATACAGAACCTCGAATATTTCTTCTGTGA 1907
QY 557 nGlnSerProValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSe 577
DB 1908 GAGGATCCAAAGCCTCAGTTCATTAACATTCATGATGACAGACTTCTCCCAATACATC 1967
QY 577 rProPheGlnAlaGlnLeuThrAspSerAspIleTrpThrAlaGluValAsnGl 597
DB 1968 TCCTCTTACAGCAGAACTAACACACGGCGAGTGGCCAACTGGACCATTCAGTACACGA 2027
QY 597 u---GluGlyAspThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrrAs 616
DB 2028 CCAACCCAAAGATCTATCATTTTGAAGCAAGATGGCCTTAGAGTGGGTAGCTACAA 2087
QY 616 pValHisLeuSerLeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaTh 636
DB 2088 AATCAATCTCAAGCTCATGATAACCAAGATAAAGACCAAGTGACCATTTAGAGTGTC 2147
QY 636 rValCysAspCysHisGlyHisValGluThrCys-----ProGlyProTrpLysGlyGl 654
DB 2148 CTTGTGTGACTGTGAAGGGCGCGCGGCTCTGTAGGAAGGCACAGCTCTGGAAGCAGG 2207
QY 654 yPhe-----IleLeuProValLeuGlyAlaValLeuAlaLeuLeuPheLeuLe 670
DB 2208 ATTGCATAATCTCGCCATCTGGGATTTCTGGAGGAATTTCTTGTGTTGTAATTCGAT 2267
QY 670 uLeuValLeuLeuLeuValArgLysLysArgLysLysLysGluProLeuLeuLeuPr 690
DB 2268 TCTGTGCTTGTCTTGTCTTTCGGAGGAGAGCGGTGGTCAAGAGAGCCCTTACTGCCCCC 2327
QY 690 oGluAspAspThrArgAspAsnValPheTrpTrpGlyGluGlyGlyGlyGluGluAs 710
DB 2328 AGAGGATGACACCCGGGACACAGTTTATCTATGATGAAGAGAGCGGAGGAGAGGA 2387
QY 710 pGlnAspTrpAspIleThrGlnLeuHisArgGlyLeuGluAlaArgProGluValVal 730
DB 2388 CCAGGACTTGTACTTGTAGCCAGCTGCACAGGGCGCTGACGCTCGGCCCTGAAGTGACT-- 2445
QY 730 uArgAsnAspValAlaProThrIleLeuProThrProMetTyrrArgProAspProAlaAs 750
DB 2446 -CGTAACAGAGTGTGACCAACCTCATGAGTGTCCCGCGGTATCTTCCCGCCCTGCCAA 2504
QY 750 nProAspGluIleGlyAsnPheIleIleGluAsnLeuLysAlaAlaAsnThrAspProTh 770
DB 2505 TCCCGATGAAATTTGAAATTTTATGATGAAATCTGAAAGCGGCTGATACTGACCCAC 2564
QY 770 rAlaProProTyrrAspThrLeuLeuValPheAspTrpGluGlySerGlySerAspAlaAl 790
DB 2565 AGCCCCGCTTATGATTTCTGTCTGTGTTTGTGACTATGAAGAGAGCGGTTCCGAGCTGC 2624
QY 790 aSerLeuSerSerLeuThrSerSerAlaSerAspGlnAspGlnAspTyrrAspTyrrLeuAs 810
DB 2625 TAGTCTGAGCTCCCTGAACTCCTCAGAGTCTAGCAAAAGACAGGACTATGACTACTTGA 2684
QY 810 nGluTrpGlySerArgPheLysLysLeuAlaAspMetTyrrGlyGlyGlyGluAsp 828

QY 298 eGlnAlaThrAspMetAspGlyAspGlySerThrThrAlaValAlaValGluI 318
DB 1131 TCAAGCTGCTGACCTTCAAGCTGAGGGGTTAAGCACACACACACAGCTGTGACACAT 1190
QY 318 eLeuAspAlaAsnAspAlaProMetPheAspProGlnIlystYrGluAlaHisValPr 338
DB 1191 CACTGACACCAACGATAATCTCGATCTTCAATCCACACGACGACAGGGTCAAGTGCC 1250
QY 338 oGluAsnAlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAs 358
DB 1251 TGAGAACGAGGCTACGTCGTAATCAACACACACTGAAGTGACTGATGTCGATGCCCAA 1310
QY 358 nSerProAlaThrArgAlaThrThrLeuIleMetGlyGlyAspAspGlyAspHisPheTh 378
DB 1311 TACCCACGCTGGGAGGCTGTATACACCATATG---AATGATGATGGTGACAAATTTGT 1367
QY 378 rIleThrThrHisProGluSerAsnGlnGlyIleLeuThrThrArgIlySGlyLeuAspPh 398
DB 1368 CGTCACCAACATCAAGTGAACACGATGGCATTTTGAAPACACAAAGGCTTGATTT 1427
QY 398 eGluAlaIlyAsnGlnHisThrLeuTyValGluValThrAsnGluAlaProPheValle 418
DB 1428 TGAGGCCAAGCAGCAGTACATTTACAGCTAGCAGTACGAAATGTGGTACCTTTTGAGGT 1487
QY 418 uIlyLeuProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaPr 438
DB 1488 CTCCTCTACACCTCCACAGCCCGCTGACCGTGGATGTGTGATGTGAATGAAGGCC 1547
QY 438 oValPheValProProSerIlyValValGluValGlnGlyIleProThrGlyGluPr 458
DB 1548 CATCTTTGTGCTCTGCTGAAAGAGAGAGTGAAGTGCAGGACTTTGGCGTGGCCACGA 1607
QY 458 oValCysValTyThrAlaGluAspProAsPlys---GluAsnGlnIlystYrAr 477
DB 1608 AATCATCTCTACATCTGCCAGGAGCCAGACATTTATGAAACAGAAATAACATATCG 1667
QY 477 gIleLeuArgAspProAlaGlyTyTrpLeuAlaMetAspProAspSerGlyGlnValThrAl 497
DB 1668 GATTTGGAGACACACTCCCACTGGCTGGAGATTAATCCGACACACTGGTGCATTTCCAC 1727
QY 497 aValGlyThrLeuAspArgGluAspGluGlnPheValArgAsnIleTyGluValme 517
DB 1728 TCGGCTGAGCTGGACAGGAGGATTTTGGACACGTGAAGAACACGACGATACACAGCCCT 1787
QY 517 tValLeuAlaMetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuTh 537
DB 1788 AATCATAGCTACAGACATGTTCTCCAGTTGCTACTGGAACAGGACACTTCTGTGTAT 1847
QY 537 rLeuIleAspValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAs 557
DB 1848 CTTGTCTGATGTGAATGACACGCCCCCATACAGAACCTCGAACTATATTTCTTCTGTGA 1907
QY 557 nGlnSerProValArgHisValLeuAsnIleThrAspIlyAspLeuSerProHisThrSe 577
DB 1908 GAGGAATCAAGCTCAGTCAATTAATTCATGATGACACCTTCTCCCATATATC 1967
QY 577 rProPheGlnAlaGlnLeuThrAspAspSerAspIleTyTrpThrAlaGluValAsnGl 597
DB 1968 TCCCTTCACAGCAGAACTAAACACACGCGGAGTGCACCACTGGACCAATTCAGTACAACGA 2027
QY 597 u---GluGlyAspThrValValLeuSerLeuIlystYrPheLeuIlySGlnAspThrTyAs 616
DB 2028 CCCAACCCCAAGAACTATCATTTTGAAGCCAAAGATGCGCTTAGAGTGGGTGACTACAA 2087
QY 616 pValHisLeuSerLeuSerAspHisGlyAsnIlySGlnLeuThrValIleArgAlaTh 636
DB 2088 AATCAATCTCAAGCTCATGATTAACACAGATAAAGACCAAGTACCACCTTAGAGGTGAG 2147
QY 636 rValCysAspCysHisGlyHisValGluThrCys-----ProGlyProTyPlysGlyI 654
DB 2148 CGTGTGTGACTGTGAAGGGCGCGCGCTGTGTAGGAAGCAGCAGCTGTGGAAGCAGG 2207

QY 654 yPhe-----IleLeuProValIleGlyAlaValLeuAlaLeuLeuPheLeuLe 670
DB 2208 ATTGCAAAATCTGTCATTTCTGGGATTTCTTGGAGGAATTTCTGTTTGTCTAATCTGAT 2267
QY 670 uLeuValLeuLeuLeuValArgIlystYrArgIlySGlnIlySGluProLeuLeuLeuPr 690
DB 2268 TCTGCTGCTCTTGTCTTCTTCTCGGAGGAGCGGTGTCAAAGAGCCTTACTTGCCTCC 2327
QY 690 oGluLeuAspThrArgAspAsnValPheTyTyGlyGluGluGlyGlyGlyGluAs 710
DB 2328 AGAGGATGACACCCGCGACACGTTTATCTATCTATGATGAAGAGGAGCGGAGAGAGA 2387
QY 710 pGlnAspTyAspIleThrGlnLeuHisArgIlyLeuGluAlaArgProGluValValle 730
DB 2388 CAGAGCTTTGACTTGGCCAGCTGCACAGGGCTCGACGCTCGCTGAGTGAAGTACT- 2445
QY 730 uArgAsnAspValAlaProThrIleIleProThrProMetTyArgProArgProAlaAs 750
DB 2446 -CGTAACGACGTTGACCAACCTCATGAGTGTCCCGGTATCTTCCCGCTGCCAA 2504
QY 750 nProAspGluIleGlyAsnPheIleIleGluAsnLeuIlystYrAlaAsnThrAspProTh 770
DB 2505 TCCGATGAATTTGAATTTTATGATGAATCTGAAGCGGCTGATCTGACCTGACCCAC 2564
QY 770 rAlaProProTyAspThrLeuLeuValPheAspTyGluGlySerGlySerAspAlaAl 790
DB 2565 AGCCCGCTTATGATTTCTGCTGTTGACTATGAAAGAGCGGTCCGAGCTGC 2624
QY 790 aSerIleuSerSerLeuThrSerAlaSerAspGlnAspGlnAspTyAspTyLeuAs 810
DB 2625 TAGTCTGACTCCCTGAACTCTCAGAGTCAGACAAAGACCAAGGACTATGACTTCTGAA 2684
QY 810 nGluTrpGlySerArgPheIlyLeuAlaAspMetTyGlyGlyGlyGluAsp 828
DB 2685 CGAATGGGCAATCGTTCAAGAGCTGGCTGACATGTACGAGGCGCGAGGAC 2739
RESULT 4
US-09-798-267-1
; Sequence 1, Application US/09798267
; Patent No. 6406870
; GENERAL INFORMATION:
; APPLICANT: Brenner, Michael
; APPLICANT: Ceppek, Karyn
; TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic E-cadherin
; FILE REFERENCE: L0560/7008RP
; CURRENT APPLICATION NUMBER: US/09/798,267
; CURRENT FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: US 08/237,919
; PRIOR FILING DATE: 1994-05-03
; PRIOR APPLICATION NUMBER: PCT/US 95/05518
; PRIOR FILING DATE: 1995-05-03
; PRIOR APPLICATION NUMBER: US 08/732,429
; PRIOR FILING DATE: 1996-11-01
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2808
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (109)..(2742)
; NAME/KEY: mat_peptide
; LOCATION: (2740)..()
; NAME/KEY: misc_feature
; LOCATION: (800)..(808)
; OTHER INFORMATION: HAV tripeptide
; NAME/KEY: misc_feature
; LOCATION: (2225)..(2295)
; OTHER INFORMATION: transmembrane
; NAME/KEY: misc_structure
; LOCATION: (2296)..(2746)

951	Db	AGCCACACAGACGGGACGATGATGGAACACCTACAATCGCCGCCATCGCTTACACCATCCT	101
258	Qy	sSerGlnGluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyTh	278
1011	Db	CAGCCAAGATCCTTGAGCTCCCTGACAAAATATGTTACCATTAACAGGAACACAGGAGT	1070
278	Qy	rIleSerValIleSerSerGlyLeuAspArgGluLeuValProGluTyrThrLeuThrIle	298
1071	Db	CATCAGTGTGGTCACCACTCGGTGGACCGAGAGATTTCCTACGTATACCTCGTGGT	1130
298	Qy	eGlnAlaThrAspMetAspGlyAspGlySerThrThrAlaValAlaValValGluIle	318
1131	Db	TCAGCTGCTGCCTTCAAGTCAAGGTGAGGTAAAGCACAACAGCAACAGCTGTGATCAGT	1190
318	Qy	eLeuAspAlaAspAspAlaProMetPheAspProGlnLysTyrGluAlaHisValPr	338
1191	Db	CACTGACACCAACGATTAATCCTCCGATCTTCAATCCACCCACGCTACAAGGTCAGGTGCC	1250
338	Qy	oGluAsnAlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAs	358
1251	Db	TGAGAACGAGGCTAACGTCGTATATCACCACTGAAGATGACTGATGCTGATGCCCCCAA	1310
358	Qy	nSerProAlaIleArgAlaThrTyrIleuIleMetGlyGlyAspAspGlyAspHisPheTh	378
1311	Db	TACCCAGCGTCGGAGGCTGTATACACCATATTG--AATGATGATGGTGGCAATTTGT	1367
378	Qy	rIleThrThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPh	398
1368	Db	CGTACACACAAATCCAGTGAACCAACGATGCGATTTTGAAGAACGAAAGGGCTTGATTT	1427
398	Qy	eGluAlaLysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValle	418
1428	Db	TGAGGCCAAGCAGCAGTACATCTACACGTAGCAGTACGAATGTGGTACCTTTTGAGGT	1487
418	Qy	uLysLeuProThrSerThrAlaThrIleValHisValGluAspValAsnGluAlaPr	438
1488	Db	CTCTCTCACACCTCCACAGCCACCGCTCACCGTGGATGTCTGGATGTGAATGAAGCCCC	1547
438	Qy	oValPheValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluPr	458
1548	Db	CATCTTTGTGCTTCCTGAAAGAGAGTGAAGTGTCCGAGGACTTGGCGTGGCCGACGA	1607
458	Qy	oValCysValTyrThrAlaGluAspProAspLys---GluAsnGlnLysIleSerTyrAr	477
1608	Db	AATCAATCCTCATCCTGCGCAGGAGCAGACATTTATGGAACAGAAAAATAACATATCG	1667
477	Qy	gIleLeuArgAspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAl	497
1668	Db	GAITTCGAGAGACATCGCAACTGGCTGGAGATTATCCGACACTGGTGCCATTTCAC	1727
497	Qy	aValGlyThrLeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMe	517
1728	Db	TCGGGCTGAGCTCGACAGGGAGATTTTGAGCAGCTGGAAGAACAGCACGCTACACGCCCT	1787
517	Qy	vValLeuAlaMetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuTh	537
1788	Db	AATCATAGCTACAGACAAATGGTTCCTCCAGTTCCTACTGTGAACAGGAGACATTTCTGTGAT	1847
537	Qy	rLeuIleAspValAsnAspHisGlyProValProGluProAspGlnIleThrIleCysAs	557
1848	Db	CCGTGCTGATGTGAATGACAACGCCCCCATACCAACCTCGAACTATATTTCTCTGTGA	1907
557	Qy	nGlnSerProValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSe	577
1908	Db	GAGGAATCCAAAGCCTCAGGTCATAAACATTCATGATGCAGACCTTCCTCCCAATACATC	1967
577	Qy	rProPheGlnAlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGl	597
1968	Db	TCCTTTCACAGCAGAACTAACACGGGGAGTGGCCAACTGGACCATTCATGACACAGA	2027
597	Qy	u---GluGlyAspThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAs	616
2028	Db	CCCAACCCCAAGAAATCTATCATTTTGAAGCAAGATGGCTTTAGAGTGGGGTACTACAA	2087

REGISTRATION NUMBER: 36,637
REFERENCE/DOCKET NUMBER: B0801/7023

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2908 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: liver

FEATURES:

NAME/KEY: CDS
LOCATION: 109..2745
FEATURE: sig_peptide
LOCATION: 109..558
FEATURE: mat_peptide
LOCATION: 559..2742
FEATURE: cytoplasmic tail
LOCATION: 2296..2746
FEATURE: HAV tripeptide
LOCATION: 800..808
FEATURE: transmembrane
LOCATION: 2225..2295

PCT-US95-05518-1

Alignment Scores:

Pred. No.: Length: 2808
Score: 2349.50 Matches: 474
Percent Similarity: 67.24% Conservativity: 117
Best Local Similarity: 53.92% Mismatches: 226
Query Match: 53.78% Indels: 62
DB: Gaps: 14

US-09-916-849A-1 (1-829) x PCT-US95-05518-1 (1-2808)

QY	Dy	ProArgGlyProLeuAlaserLeuValCysTyrPheThrValLeuGlnCysAla-Al	23
Db	129	CCTCTCGGGCCGTGCCTTCCTCT-----TGCTCTCAGACGCCGAG	182
QY	23	aSerGluProCysArgAlaValPheArgGluAlaGlutThrLeuGluAlaGlyGlyAl	43
Db	183	CCCCCTCTCTCTCT-----TTGACGCGAGACTCACGTCGCCGCGGCCGCCACC	236
QY	43	agLuGlnGluProGlyGlnAlaLeuGlyLysVal---PheMetGlyCysProGlyGlnGI	62
Db	237	TGAAGAAGGCCG-----CGTCTGGGCGAGATGAATTITCAAGATTGCACCGGTGACA	290
QY	62	u-----	62
Db	291	AAGGACAGCATTTTTCTCTCACACCGATTCCGAAAGTGGGCACAGATGGTGTGATTACAGT	350
QY	63	-----ProAlaLeuPheSerThrasAsnAspPheThrValArgAsnGlyGluTh	80
Db	351	CAAAGGCCCTTACGGTTTCATAACCCAACAGATCCATTCTTGGTCTACGCTGGGACTC	410
QY	80	rValGlnGluArgSerLeuLys-----GluArgAs	91
Db	411	CACCTACAAAAGTTTTCACCAAAAGTCACGCTGAATACAGTAGTGGGCGAACCCAGCCC	470
QY	91	npro-----LeuLysIlePhePro--Se	98

pValHisLeuSerLeuSerHisGlyAsnLysGluGlnLeuThrValileArgAlath 636
: : : : :
2088 AATCAATCTCAAGCTCATGNATACCAAGTAACAGACCAAGTACCACTTAGAGGTGAG 2147
: : : : :
636 rValCysaspCysHisgLyHisValgluthrCys-----ProGlyProTrpLysglyGI 654
: : : : :
2148 CGTGTGTGACTGTGAAGCGCGCGCGGCTGTGTAGGAAGGCGACAGCTGTGCAAGCAGG 2207
: : : : :
654 yphe-----lleLeuProValLeuGlyAlavalLeuAlalaLeuHeule 670
: : : : :
2208 ATTGCAAAATTCCTGCCATCTCGGGATCTTGGAGGAATCTTCTTGTGTGTAATTTGAT 2267
: : : : :
670 uLeuValLeuLeuLeuValArGLysLysArgLyslleyLsgluProLeuLeuPr 690
: : : : :
2268 TCTGCTGCTCTTGTCTTCTTCGAGGAGAGCGGTGTCAAAGAGCCCTTACTTGCCTCC 2327
: : : : :
690 oGiUasppThrArgaspAnvalPhetyTyrglyGlucluglyGlyGluGuas 710
: : : : :
2328 AGAGATGACACCOGGGCAACGTTTATCTATGATGAAGAAGAGCGGCGAGAGGA 2387
: : : : :
710 pGlnAspyrAspilethrGlnLeuHisArgGlyLeuGluLaalargProGluValVall 730
: : : : :
2388 CCNAGACITTGACITGAGCCAGCTGCACAGGGCTCGAGCTCGGCTGAAGTGACT-- 2445
: : : : :
730 uArgasnspvallalpProThrilleleProThrPrometTyraRgpProargProlaas 750
: : : : :
2446 -CGTAACGACGTGTGACCAACCCCTCATGAGTGTCCCCCGGTATCTCCCCGCGCTGCCAA 2504
: : : : :
750 nProaspGluileGelivAnsphelleillegluasnLeuLysAlaalasanThrapproth 770
: : : : :
2505 TCOCGATGAANTGAAAAATTTTATGATGAANAATCTGAAGCGGCTGATACTGACCCAC 2564
: : : : :
770 rAlaprophryrAspthrLeuValphaSpTYrCluglyserglySeraspalaal 790
: : : : :
2565 AGCCCCGGCTATGATTCTCTCTGCTGTGTTGACTATGAGAAGAACGGTTCCGAAGCTGC 2624
: : : : :
790 aserLeuSerleruThrserralaraserglinAspnAspyrAsptvrlauas 810
: : : : :
2625 TAGTCTGAGCTCCCTGAACCTCTGAGACTGACAAAGACCGAGACTATGACTACTTGAA 2684
: : : : :
810 ngLutrpglyseryarXpghelysLeuAlaaSPmettyrGlyGlyGluusp 828
: : : : :
2685 CGAATGGGCAATCGTCTCAAGAACTGGCTGACATGTACGAGGCGCGGAGGAC 2739

RESULT 5

PCT-US95-05518-1

Sequence 1, Application PC/TUS9505518

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic E-cadherin Interactions with T Lymphocytes

NATURE OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSER: Wolf, Greenfield & Sacks P.C.

STREET: 600 Atlantic Avenue

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent In Release #1.0, Version #1.25

SOURCE APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/05518

FILING DATE: herewith

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/237,919

FILING DATE: 3 May 1994

ATTORNEY/AGENT INFORMATION:

NAME: Plumer, Elizabeth R.

Db 471 CCGGCCCATCAGGCTCGTTTCTGGAAATCCAAAGCAGAAATGCTCACATTTTCCCAACTC 530
Qy 98 rlyAsrGileuAArgHisIysArgAspIrpValValAlaProIleSerValProgl 118
Db 531 CTTCTCTGCTCAGAGACAGAGAGACTGGGTATTCTCCCATCAGCTGCCCA 590
Qy 118 uAsnGlySgIyProPheProGlnArgLeuAenGlnLeuLysSerAsnLysAspArgAs 138
Db 591 AAATGAAAGGCCCATTTCTTAAACCTCGTTTCTAGATCAATCCAAACAAAGACAAGA 650
Qy 138 pThrIlySerPheThrSerIleThrGlyProGlyAlaAspSerProGluGlyValph 158
Db 651 AGCAGAGTTTCTCAGCATCCTGCGCCAGAGCTGACACACCCCTGTGTGTCTT 710
Qy 158 eAlaValGluSgIyThrGlyIrpLeuLeuLeuAsnLysProLeuAspArgGluGluI 178
Db 711 TATTATTGAAAGAGAAACAGGATGCTGAGGTGACAGAGCTCTGGATAGAGACGCA 770
Qy 178 eAlaLysTyIrpLeuPheGlyHisAlaValSerGluAenGlyAlaSerValGluAspPr 198
Db 771 TGCCACATACACTCTCTCTCTCAGCTGTGTCTCAACCGGAATGCAATGAGGATCC 830
Qy 198 oMetAsnIleSerIleIleValThrAspGlnAsnAspHisIlyProLysPheThrGlnAs 218
Db 831 AATGAGATTTGATCACGGTAACCGATCAGATGACAAAGCCCGCAATTCACCCAGGA 890
Qy 218 pThrPheArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValTh 238
Db 891 GGTCTTTAAGGGGTCTGTATGGAAGGTGCTCTTCCAGGAACCTCTGTGTGAGGTGAC 950
Qy 238 rAlaThrAspGluAspAlaIleTyIrpThrTyIrpAsnGlyValValAlaTyIrpSerIleHi 258
Db 951 AGCCACAGACGGGACGATGATGACCACTCAATGCCGCCATCGCTTACACCATCTCT 1010
Qy 258 sSerGlnGluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyTh 278
Db 1011 CAGCCAGATCTGAGTCCCTGTGACAAATAATGTTCCACATTAACAGGAACACAGAGT 1070
Qy 278 rIleSerValIleSerSerGlyLeuAspArgGluLysValProGluTyIrpThrLeuThrI 298
Db 1071 CATCAGTGTGTCACCTGCGGTGACCGGAGAGATTTCCCTACGTATACCTCTGTGTGT 1130
Qy 298 eGlnAlaThrAspMetAspGlyAspGlySerThrThrAlaValAlaValValGluI 318
Db 1131 TCAAGCTGCTGACCTTCAAGGTGAGGGTTAAGCACACACAGCACACCTGTGTATCAGT 1190
Qy 318 eLeuAspAlaAsnAspAsnAlaProMetPheAspProGlnLysTyIrpGluAlaHisValPr 338
Db 1191 CACTGACACCAACAGATAATCTCCGATCTTCAATCCACACACAGTACAAAGGTTCAGGTGCC 1250
Qy 338 oGluAsnAlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAs 358
Db 1251 TGAGAACGAGCTAACCTCGTAATCACCACTGAAAGTGACTGTGCTGATGCCCCCAA 1310
Qy 358 nSerProAlaTrpArgAlaThrTyIrpLeuIleMetGlyGlyAspAspGlyAspHisPheTh 378
Db 1311 TACCCAGCTGGGAGCTGTATACACATATTG---AATGATGATGGTGACAAATTTGT 1367
Qy 378 rIleThrThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPh 398
Db 1368 CGTCACCAAAATCCATGTAACACAGATGGCATTTTGAACACAGCAAGGCGCTTGATTT 1427
Qy 398 eGluAlaLysAsnGlnHisThrLeuTyIrpValGluValThrAsnGluAlaProPheValle 418
Db 1428 TGAGGCCAACGACGATACATTCTACAGTAGCAGTGACGAATGTGGTACCTTTTGAGGT 1487
Qy 418 uLysLeuProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaPr 438
Db 1488 CTCTCTCACACCTCCACACCCACCGTACCGTGGATGTCTGGATGTGAATGAAGGCC 1547
Qy 438 oValPheValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluPr 458
Db 1548 CATCTTGTGCTTCTTGAAAGAGAGTGGAGTGTCCGAGGACTTTTGGCGGTGGGCCAGGA 1607

Qy 458 oValCysValTyIrpThrAlaGluAspProAspLys--GluAsnGlnLysIleSerTyIrpAr 477
Db 1608 AATCACATCTCTACACTGCCAGGAGCCAGACACATTTATGAAACAGAAATAACATATCG 1667
Qy 477 gIleLeuArgAspProAlaGlyIrpLeuAlaMetAspProAspSerGlyGlnValThrAl 497
Db 1668 GATTGGAGAGACACTCGCAACTGGCTGGAGATTAAATCCGAGACACTGGTCCCATTTCCAC 1727
Qy 497 aValGlyThrLeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyIrpGluValMe 517
Db 1728 TCGGGCTGAGCTGGACAGGAGGATTTTGACGACGTCGTGAAGAACAGACGCTACACAGCCCT 1787
Qy 517 tValLeuAlaMetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuTh 537
Db 1788 AATCATAGCTACAGACAATGGTTCTCCAGTTGCTACTGGAACAGGAGACACTTCTGTGTAT 1847
Qy 537 rLeuIleAspValAsnAspHisGlyProValProGluProArgGlnIleThrIleThrIleCysAs 557
Db 1848 CCTGTCTGATGTGAATGACAAACGCCCCCATACAGAACCTCGAACTATATTCTTCTGTGA 1907
Qy 557 nGlnSerProValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrse 577
Db 1908 GAGGAATCCAAAGCCTCAGTCTCAATAACATTCATGATGCAGACCTTCTCTCCAATACATC 1967
Qy 577 rProPheGlnAlaGlnLeuThrAspAspSerAspIleTyIrpThrAlaGluValAsnGl 597
Db 1968 TCCCTTCACAGCAGAACTAACACACGGCGAGTGCCTTACGACCTTCTAGAGTGGGTGACTACAA 2027
Qy 597 u----GluGlyAspThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyIrpAs 616
Db 2028 CCAACCCCAAGATCTATCATTTTGAAGCCAAAGATGGCCTTAGAGTGGGTGACTACAA 2087
Qy 616 pValHisLeuSerLeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaTh 636
Db 2088 AATCAATCTCAAGCTCATGATAACAGAAATAAGACCAAGTACCACTTAGAGGTGAG 2147
Qy 636 rValCysAspCysHisGlyHisValGluThrCys-----ProGlyProTyIrpLysGlyI 654
Db 2148 CGTGTGTGCTGTGAAGGGGCCCGCGGCTGTAGGAAGGCACACGCTGTGCAAGCAGG 2207
Qy 654 yPhe-----IleLeuProValLeuGlyAlaValLeuAlaLeuLeuPheLeu 670
Db 2208 ATTGCAAAATCTCGTGCATTTCTGGGATTTCTGGAGGAATTTCTTGTCTTCTTAATCTGAT 2267
Qy 670 uLeuValLeuLeuLeuValArgLysLysArgLysIleLysGluProLeuLeuLeuPr 690
Db 2268 TCTGTGCTCTTCTGTCTTCTTCTCGAGGAGACGCGGTGTCAAAGAGCCCTTACTGCCCCC 2327
Qy 690 oGluAspAspThrArgAspAsnValPheTyIrpTyIrpGlyGluGlyGlyGlyGluGluAs 710
Db 2328 AGAGGATGACACCCGGGACACGTTTATTATCTATGATGAAGAAGGAGCGGAGAGAGA 2387
Qy 710 pGlnAspTyIrpAspIleThrGlnLeuHisArgGlyLeuGluAlaArgProGluValValle 730
Db 2388 CAGAGACTTTGACTTGACCCAGCTGCACAGGGGCTGAGCGCTCGGCTCGAAGTACT-- 2445
Qy 730 uArgAsnAspValAlaProThrIleProThrProMetTyIrpArgProArgProAlaAs 750
Db 2446 -CGTAACGACGTTGCACCAACCTCATGAGTGTCCCCCGGTATCTTCCCCCGCTGCCAA 2504
Qy 750 nProAspGluIleGlyAsnPheIleIleGluAsnLeuLysAlaAlaAsnThrAspProTh 770
Db 2505 TCCCGATGAATTTGAAATTTTATTGATGAATACTGAAGCGGCTGATCTGACCCCA 2564
Qy 770 rAlaProProTyIrpAspThrLeuLeuValPheAspTyIrpGluGlySerGlySerAspAlaAl 790
Db 2565 AGCCCCGCTTATGATCTCTGCTGTGTTGACTATGAAGAAGCGGTTCGAAGCTGC 2624
Qy 790 sSerLeuSerSerLeuThrSerSerAlaSerAspGlnAspGlnAspTyIrpAspTyIrpLeuAs 810
Db 2625 TAGTCTGAGCTCCCTGAACTCTCAGAGTCAAGAAACAGACAGGACTATGACTACTTGA 2684

QY 350 ThrValThrAspLeuAspAlaProAsnSerProAlaThrArgAlaThrThrLeuIleMet 369
DB 1254 ACGTGATGACCGAGATACGCCCACTCTCCAACTCGAATGCCGTTTACCGCATCATC 1313
QY 370 GlyGlyAspAspGlyAspHisPheThrThrHisProGluSerAsnGlnGlyIle 389
DB 1314 AGTGGGATCCATCCGGGCACTTACGCTCGCACAGACCCCGTAACCAACGAGGCGCATG 1373
QY 390 LeuThrThrArgLysGlyLeuAspPheGluAlaLysAsnGlnHisThrLeuThrValGlu 409
DB 1374 GTACCGTGGTGAAGCGATCGATACGAGCTCAACAGAGTTCATGCTCAGCATGATG 1433
QY 410 ValThrAsnGluAlaProPheValLeuLysLeuPro-----ThrSerThrAlaThr 426
DB 1434 GTGTCCAAACGAGGCGCCCTGGCCAGCGGAATCCAGATGCTCTTCCAGTCCACGCGAGGG 1493
QY 427 IleValValHisValGluAspValAsnGluAlaProValPheValProProSerLysVal 446
DB 1494 GTGACCATCTCCATCATCGATCAACAGAGCTCCCTACTTCCCTCAAAACCAACGAGCTG 1553
QY 447 ValGluValGlnGluGlyIleProThrThrGlyGluProValCysValThrThrAlaGluAsp 466
DB 1554 ATCCGCTGGAGAGGGGTGCCCCCGGACCGCTGACCATCGTTTACGTGTGGAC 1613
QY 467 ProAspLys-----GluAsnGlnLysIleSerTyrArgIleLeuArgAspProAlaGlyTrp 485
DB 1614 CCTACCGGTTTCATGACAGCGCTGTGAGATCTCAAAAGCTGTCCAGACCCAGCGAGCTGG 1673
QY 486 LeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThrLeuAspArgGluAsp 505
DB 1674 CTGCATCAATGCCAACCGGCGATCAACAGGTCGAGCTGCGGAGTGTGGACCGGAG--- 1730
QY 506 GluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAlaMetAspAsnGlySer 525
DB 1731 TCCCTCTACACCAAAACAACTGTACGAGGCCACCTTCTGGCGAGCTGACAAATGGGATA 1790
QY 526 ProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAspValAsnAspHisGly 545
DB 1791 CCCCCGGCCAGCGGACCGGACCCCTCAGATCTATCTGATTCATCAACGACCAACGCC 1850
QY 546 ProValProGluProArgGlnIleThrIleCysAsnGlnSerProValArgHisValLeu 565
DB 1851 CCTGAGCTGTGCCCAAGGAGCGCAGATCTGC---GAGAGGCCCACTGGAAGCCATC 1907
QY 566 AsnIleThr-----AspLysAspLeuSerProHisThrSerProPheGlnAlaGlnLeu 583
DB 1908 AACATCAGCGCGCGCGAGCTGACGTGACGCCCAACATCGGCCCTACGTCTTCGAGCTG 1967
QY 584 ThrAspAspSerAspIleTyrThrThrAlaGluValAsnGlu----- 597
DB 1968 CCC-----TTGTCCCGCGCGCGTGGGAGAACTGGACCATCACCCGC 2012
QY 598 ---GluGlyAspThrValValLeuSerLeuLys---LysPheLeuLysGlnAspThrTyr 615
DB 2013 CTGAACGGTGACTATGCCAACTCAGCTTGGCATCTGTACCTGGAGCGCGGATGTAT 2072
QY 616 AspValHisLeuSerLeuSerAspHisGlyAsnLys-----GluGlnLeuThrValIle 633
DB 2073 GACGTCCCATCATCTGCACAGACTCTGGAACCCCTCCCTGTCACACAGCTCCATCATC 2132
QY 634 ArgAlaThrValCysAspCys-----HisGlyHisValGluThrCysProGlyProTyr 651
DB 2133 AAAGTCAAGTGTGCCCATGTGTGACACAGGAGCTGCACCATCATCTGGCGGAGTGGCA 2192
QY 652 LysGlyGlyPheIleLeuProValLeuGlyAlaValLeuAlaLeuPheLeuLeuLeu 671
DB 2193 GCGGCTGGTCTG-----GGCACCGGTGCCATCGTGGCCATCTCATCTGCATCTC 2243
QY 672 ValLeuLeuLeuLeuVal-----ArgLysLysArg 681
DB 2244 ATCTGTCTGACCATGCTCTGCTGTGTGTGATGAGCGGAGAGAGGAGCGC 2303
QY 682 LysIleLysGluProLeuLeuLeuProGluAspThrArgAspAsnValPheTyrTyr 701

DB 2304 CACACGAGCAGCTGCTCATTTACCCCGGAGGACGAGCTCCGCGAAAGATCTCTCAAGTAT 2363
QY 702 GlyGluGluGlyGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisArgGly 721
DB 2364 GACGAGGAGCGGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2423
QY 722 LeuGluAlaArgProGluVal-----ValLeuArgAsnAspValAla 735
DB 2424 ---GAGACCCATGGGCGACGTCGCAAAAGCCCTTGGCGTGCCTGCGTGGATGAGCGG 2480
QY 736 ProThrIleIleProThrProMetTyr-----ArgProArgProAlaAsnProAspGlu 753
DB 2481 CCG---GTGGGCGCTGAGCCCGGAGTACCGATCAGGCCCCATGTCGCGCGGAGCGGAC 2537
QY 754 IleGlyAsnPheIleIleGluAsnLeuLysAlaAlaAsnThrAspProThrAlaProPro 773
DB 2538 ATCGGTGACTTCATCAATGAGGAGCTCCGCGTGTGTGACAAAGCAGCCCGGAGCGGCG 2597
QY 774 TyrAspThrLeuLeuValPheAspTyrGlyGlySerGlySerAspAlaAlaSerLeuSer 793
DB 2598 TATGACTTCCCTGCTGCTGCTTTCGACTACGAGGAGGAGGAGGAGGAGGAGGAGGAG 2657
QY 794 SerLeuThrSerSerAlaSerAspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGly 813
DB 2658 TCCTCAACTCATCCAGTTCGCG---GACCAAGACTACGATTACCTCAACGACTGGGCG 2714
QY 814 SerArgPheLysLysLeuAlaAspMetTyrGlyGlyGlyGluAspAsp 829
DB 2715 CCAGATTCAAGAAGCTGGCGACATGTATGAGGTGTGAAGAGGAT 2762

RESULT 7

US-08-332-643-41
; Sequence 41, Application US/08332643
; Patent No. 5639634
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Bicknell
; STREET: Two First National Plaza, 20 South Clark
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/332,643
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/872,643
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5639634and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/30795
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 346-5750
; TELEFAX: (312) 984-9740
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3048 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

TOPOLOGY: linear		
MOLECULE TYPE: CDNA		
US-08-332-643-41		
Alignment Scores:		
Pred. No.:	6.5e-157	Length: 3048
Score:	1621.50	Matches: 371
Percent Similarity:	54.38%	Conservative: 138
Best Local Similarity:	39.64%	Mismatches: 296
Query Match:	37.11%	Indels: 132
DB:	1	Gaps: 22
US-09-916-849A-1 (1-829) x US-08-332-643-41 (1-3048)		
QY	4	ProArgGlyProLeuAlaSerLeuLeuLeuGlnValCysTrpLeu-----19
DB	19	CGCGGGCGCGCGGTGCTC-CTTCTGCTGCTCTCGCTCCGGCGCGCTCCGGGCCCAT 77
QY	20	-----GlnCysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThr 37
DB	78	AATGAGGATCTTACAACCTAGAGACACCTGCAGGCTGGGTTCTCTGAAGATGATTACACG 137
QY	38	LeuGluAlaGlyGlyAlaGluGlnProGlyGlnAlaLeuGlyLysValPheMetGly 57
DB	138	GCATTAAATCTCCAAATATTCTAGAGGGGAAAGTACTTTCAGTCAAGTTTCAGCAGC 197
QY	58	CysProGlyGlnGluProAlaLeuPheSerThrAspAsnAspAspPheThrValArgAsn 77
DB	198	TGTGTGGGACCAAGGGGACACAATATGAGACCAACAGCATGGACTTCAAAGTTGGGGCA 257
QY	78	GlyGluThrValGlnGluArgSerLeuLysGluArgAsnProLeuLysIlePhePro 97
DB	258	GATGGCAGCTCTTCCGCCACCGGAGCTGCAG-----GTCCCC 296
QY	98	SerLysArgIle-----101
DB	297	TCCGAGAGGTGGCGTTTCCAGTGTACTGATGGGACAGCCAGACAGAGAGAAATGGAC 356
QY	101	-----101
DB	357	GCGTGTGTGCGGTGCTGCTGCGCCAGACCTCGTCCCGGCACTGTGACACACAGCCGAG 416
QY	101	-----101
DB	417	AAAGAAAGAGGTGCTGCTGTGACCCCTCTCGGCTCCGAGGACACCCCTGCTCGCG 476
QY	102	-----LeuArgArgHisLysArgAspTrpValValAla 112
DB	477	TGGCCCCAGCAGCAGAACCGCAACCGGCTGAGGGCGGCAACCGGACTGGGTCTATCCCA 536
QY	113	ProIleSerValProGluAsnGlyLysGlyProPheProGlnArgLeuAsnGlnLeuLys 132
DB	537	CCCATCAAGTCCCGGAGAACTCGCGCGGCGCTTCCCGCAGCAGCTCGTGAAGATCCGG 596
QY	133	SerAsnLysAspArgAspThrLysIlePheThrSerIleThrGlyProGlyAlaAspSer 152
DB	597	TCCGACAAAGACAATGACATCCCATCCGTTACAGCATCAGCGGAGTGGTGGCGCCAG 656
QY	153	ProProGluGlyValPheAlaValGluLysGluThrGlyTrpLeuLeuAsnLysPro 172
DB	657	CCCCCATGAGAGTCTTACAGCAATTAATCCATGTCGCGCGGAGTGTACTGACAGGCC 716
QY	173	LeuAspArgGluGluIleAlaLysTyrGluLeuPheGlyHisAlaValSerGluAsnGly 192
DB	717	ATGGACCGGAGGAGCAGCCCTTTACCACTCCGAGCCACGCTGTGGACATGAATGGC 776
QY	193	AlaSerValGluAspProMetAsnIleSerIleValThrAspGlnAsnAspHisLys 212
DB	777	AACAAGGTGGAGAACCCCATCGACTGTATCTACGTCATGCAATGAATGACAAACAC 836
QY	213	ProLysPheThrGlnAspThrPheArgGlySerValLeuGluGlyValLeuProGlyThr 232
DB	837	CCTGAGTTTATCAACACGAGTCTACAACTGCTCCGTGGAGGAGGCTCCCAAGCCAGGCACC 896
QY	233	SerValMetGlnValThrAlaThrAspGluAspAspAlaIleTyrThrTyrAsnGlyVal 252
DB	897	TACGTGATGACCATCACGGCCACAGCTGCTGACACAGACACC---ACGGCCAAACGGGATG 953
QY	253	ValAlaTyrSerIleHisSerGlnGluProLysAspProHisAspLeuMetPheThrIle 272
DB	954	GTGCGGTACCGGATCGTGACCCAGACCCACAGAGCCGCTCCAGAAATATGTTCCACATC 1013
QY	273	HisArgSerThrGlyThrIleSerValIleSerSerGlyLeuAspArgGluLysValPro 292
DB	1014	AACAGCGAGACTGGAGATATCGTCAGTGGCGGTGGTGGGACCGGAGAGAAAGTTTCAG 1073
QY	293	GluTyrThrLeuThrIleGlnAlaThrAspMetAspGlyAsp-----GlySerThr 309
DB	1074	CAGTACACAGTATCGTTTCAGGCCACAGATATGAGAGGAATCTCAACTATGGGCTCTCA 1133
QY	310	ThrThrAlaValAlaValIleLeuLeuAspAlaAsnAspAlaProMetPheAsp 329
DB	1134	AACACAGCCACAGCCATCATCAGGTGACAGATGTGAATGACAACCCGTCAGAATTTACC 1193
QY	330	ProGlnLysTyrGluAlaHisValProGluAsnAlaValGlyHisGluValGlnArgLeu 349
DB	1194	GCCAGCAGCTTTTCAGGGGAGGTCCCCGAAACACGCGTGGAGACCGTGTGCGAAACCTC 1253
QY	350	ThrValThrAspLeuAspAlaProAsnSerProAlaTrpArgAlaThrTyrLeuIleMet 369
DB	1254	ACGTTGATGGACCGAGATCAGCCCCACCTCTCCAAATCGGAATGCGGTTTACCGCATCATC 1313
QY	370	GlyGlyAspAspGlyAspHisPheThrIleThrHisProGluSerAsnGlnGlyIle 389
DB	1314	AGTGGGATCCATCCGGGCACCTTCAGCTCCGACAGACCCCGTAACCAACAGAGGGGATG 1373
QY	390	LeuThrThrArgLysGlyLeuAspPheGluAlaLysAsnGlnHisThrLeuTyrValGlu 409
DB	1374	GTACCGTGTGAGGAGTGTGACTACAGCTCAACAGAGCTTTCATGCTGACAGTGTG 1433
QY	410	ValThrAsnGluAlaProPheValLeuLysLeuPro-----ThrSerThrAlaThr 426
DB	1434	GTGTCCACACGAGCGCCCTCGCCAGCGGAATCCAGATGTCTTCCAGTCCACGCGCAGG 1493
QY	427	IleValValHisValGluAspValAsnGluAlaProValPheValProSerLysVal 446
DB	1494	GTGACATCTCCATGATGACATCAACAGGCTCTCTACTTCCCTCAACACCAAGCTG 1553
QY	447	ValGluValGlnGlyIleProThrGlyGluProValCysValTyrThrAlaGluAsp 466
DB	1554	ATCCGCTGGAGAGGCGGTGCCCCCGGCACCGTGTGACCCAGCTTTCAGCTGTGGAC 1613
QY	467	ProAspLys---GluAsnGlnLysIleSerTyrArgIleLeuArgAspProAlaGlyTrp 485
DB	1614	CCTGACCGGTTTCATGACGAGCTGTGAGATACTCAAAGCTGTGACACCCAGGAGTGG 1673
QY	486	LeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThrLeuAspArgGluAsp 505
DB	1674	CTGACATCAATGCCACCAACCGCCAGATCACACCGTGGCAGTGTGACCGTGTGAG--- 1730
QY	506	GluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAlaMetAspAsnGlySer 525
DB	1731	TCCTCTACACCAAAACCAACGTCTACGAGGCCACCTCTCTGGCAGCTGACATGGATA 1790
QY	526	ProProThrThrGlyThrGlyThrLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 545
DB	1791	CCCCCGCCAGCGGACCGGACCTTCCAGATCTATCTCATTTGACATCAACCAACGCGC 1850
QY	546	ProValProGluProArgGlnIleThrIleCysAsnGlnSerProValArgHisValLeu 565
DB	1851	CCTGAGCTGTGCGCCCAAGGAGGCGGAGATCTGC---GAGAGGCCCAACCTGAACGCCATC 1907
QY	566	AsnIleThr-----AspLysAspLeuSerProHisThrSerProPheGlnAlaGlnLeu 583
DB	1908	AACATCACGGCGCGCGAGCTGACGTGACGCCACCCCAACATCGGCCCTTACCTCTCGAGCTG 1967

QY 584 ThrAspSerAspIleThrTrpThrAlaGluValAsnGlu----- 597
Db 1968 CCC-----TTTGTCCCGCGCGCGTGGGAAGACTGACCATCACCGC 2012
QY 598 ---GluGlyAspThrValValLeuSerLeuLys---LysPheLeuLysGlnAspThrTyr 615
Db 2013 CTGAACGGTGTGATGCGCAACTGAGTGGCGATCTGCTGCTGAGCGCGGATGAT 2072
QY 616 AspValHisLeuSerLeuSerAspHisGlyAsnLys-----GluGlnLeuThrValle 633
Db 2073 GACGTCCCATCATGCTGTCAGAGACTCTGGAACCCCTCCCTGTCACACGTCCTCATC 2132
QY 634 ArgAlaThrValCysAspCys-----HisGlyHisValGluThrCysProGlyProTrp 651
Db 2133 AAAGTCAAGGTGTGCCATGTGATGACAAACGGGACTGCACCACTTGGCGAGTGGCA 2192
QY 652 LysGlyGlyPheIleLeuProValLeuGlyAlaValLeuAlaLeuPheLeuLeu 671
Db 2193 CGCGCTGCTCG-----GGCACCGGTGCCATCGTGGCCATCTCATCTGCTCCTC 2243
QY 672 ValLeuLeuLeuVal-----ArgLysLysArg 681
Db 2244 ATCTGTGTGACCATGCTCTGCTGTTCTCATGTGGATGAAGCGCGAGAGAGGAGCGC 2303
QY 682 LysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 701
Db 2304 CACACGAGCAGCTGCTCATTCACCGCGAGGACGCTCCGGAAGATCCTCAAGTAT 2363
QY 702 GlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArgGly 721
Db 2364 GACGAGGAGCGGTGGGAGGAGGACGAGCTACGACCTCAGCCAGCTGCGAGCGC 2423
QY 722 LeuGluAlaArgProGluVal-----ValLeuArgAspAspValAla 735
Db 2424 ---GAAGCATGGGCGACGTGCCAAGAAAGCCCTGGCGTGGCTGCGTGGATGAGCGG 2480
QY 736 ProThrIleIleProThrProMetTyr-----ArgProArgProAlaAsnProAspGlu 753
Db 2481 CCG---GTGGGCCCTGAGCCCGCATCCGATCAGGCGCCATGGCGCCAGCGCGAC 2537
QY 754 IleGlyAsnPheIleIleGluAsnLeuLysAlaAlaAsnThrAspProThrAlaProPro 773
Db 2538 ATCGGTGACTTCATCAATGAGGACTCGCGCTGCTGACAAACGCCACCGCACCCCCC 2597
QY 774 TyrAspThrLeuLeuValPheAspTyrGluCysSerGlySerAspAlaAlaSerLeuSer 793
Db 2598 TATGACTCCCTGCTGGTCTTCGACTACGAGGGAGCGGCTCCACCGCAGGCTCCGTCAGC 2657
QY 794 SerLeuThrSerSerAlaSerAspGlnAspTyrAspTyrAspTyrLeuAsnGluTrpGly 813
Db 2658 TCCTGTGACTCATCCAGTTCGGG---GACCAAGACTACGATTACCTCAAGCTGCGGC 2714
QY 814 SerArgPheLysLysLeuAlaAspMetTyrGlyGlyGlyGluAspAsp 829
Db 2715 CCAGATTCAAGAGCTGGCGGACATGTATGGAGGTGGTGAAGAGGAT 2762

RESULT 8

US-08-332-638-47

; Sequence 47, Application US/08332638

; Patent No. 5646250

; GENERAL INFORMATION:

; APPLICANT: Suzuki, Shintaro

; TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS

; NUMBER OF SEQUENCES: 62

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &

; ADDRESSEE: Borun

; STREET: 6300 Sears Tower, 233 S. Wacker Drive

; CITY: Chicago

; STATE: Illinois

; COUNTRY: USA

; ZIP: 60606

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/332,638
; FILING DATE: 01-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,643
; FILING DATE: 17 APR 1992
; APPLICATION NUMBER: US/08/049,460
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5646250 and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31340
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 47:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3048 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-332-638-47

Alignment Scores:
Pred. No.: 6.5e-157 Length: 3048
Score: 1621.50 Matches: 371
Percent Similarity: 54.38% Conservative: 138
Best Local Similarity: 39.64% Mismatches: 296
Query Match: 37.11% Indels: 132
DB: 1 Gaps: 22

US-09-916-849A-1 (1-825) x US-08-332-638-47 (1-3048)

QY 4 ProArgGlyProLeuAlaSerLeuLeuGlnValCysTrpLeu----- 19
Db 19 CGCGGGCGCGCGCTGCTC-CTTCTGTCTCTCGCTCTCGCGCGCGTCCGGGCCCAT 77
QY 20 -----GlnCysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThr 37
Db 78 AATGAGGATCTTACAACTAGAGAGACCTCAAGGCTGGTCTCTGAAGATGATTACAGC 137
QY 38 LeuGluAlaGlyGlyValaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGly 57
Db 138 GCATTAACTCCAAAAATTTCTAGAGGGGAAAAGCTACTTCAAGTCAAGTTCAGCAGC 197
QY 58 CysProGlyGlnGluProAlaLeuPheSerThrAspAsnAspAspPheThrValArgAsn 77
Db 198 TGTGTGGGACCAAGGGGACACAATATGAGACCAACAGCATGAGCTTCAAAGTTGGGCA 257
QY 78 GlyGluThrValGlnGluArgSerLeuLysGluArgAsnProLeuLysIlePhePro 97
Db 258 GATGGGACAGTCTTCGCCCGGAGCTGCAG-----GTCGCC 296
QY 98 SerLysArgIle----- 101
Db 297 TCCGAGCAGGTGGCTTCCCGTGTGACTGCTGAGGACCCACAGCAGCAAGATGGAC 356
QY 101 ----- 101
Db 357 GCGGTGTGCGGTGCTGTGGTGGCCCGACCTCGTCCCGCCTCTGGACACAGCGCGAG 416
QY 101 ----- 101
Db 417 AAAGGAAAGAGGTGCTGGTCTTGAGCCCTCTCCGCTCCGAGGACACACCTGCTGCGG 476
QY 102 -----LeuArgArgHisLysArgAspTrpValValAla 112

```
Db 477 TGGCCCGCAGCAGACAGCCAGCGGCTGAGCGGGCGCAACGGGACTGGGTCAATCCCA 536
Qy 113 ProIleSerValProGluAenGlyLysGlyProPheProGlnArgIleuAenGlnLeuLys 132
Db 537 CCCATCAACGTGCCGAGAACTCGCGGGGCCCTTCCCGCAGCAGCTCGTGAGAGATCCGG 596
Qy 133 SerAsnLysAspArgAspThrLysIlePheThrSerIleThrGlyProGlyAlaAspSer 152
Db 597 TCCGACAAAGACAAATGATATCCCATTCGGTACAGCATACGGGAGTGGGTGCGACCCAG 656
Qy 153 ProProGluGlyValPheAlaValGluLysGluThrGlyTrpLeuLeuLeuAenLysPro 172
Db 657 CCCCCCATGGAGGTCTTCAGCATTAATCCATGTCGCGCGCGGATGTACGTCAACAGGCC 716
Qy 173 LeuAspArgGluGluIleAlaLysTyrrGluLeuPheGlyHisAlaValSerGluAenGly 192
Db 717 ATGACCGGGAGGACGCGCTCTTACACCTCCGAGCCCGCAGCTGTGGACATGAATGGC 776
Qy 193 AlaSerValGluAspProMetAsnIleSerIleValThrAspGlnAsnAspHisLys 212
Db 777 AACAGGTGGAGAACCCCATCGACTCTACATCTACGTATCGACATGAATGACAACCCAC 836
Qy 213 ProLysPheThrGlnAspThrPheArgLysThrValLeuGluGlyValLeuProGlyThr 232
Db 837 CCTGAGTTTCATCAACACCGGTCTACAACCTGCTCGGTGACGAGGCTCCAAAGCCAGGC 896
Qy 233 SerValMetGlnValThrAlaThrAspGluAspAlaIleTyrrThrTyrrAsnGlyVal 252
Db 897 TACGTGATGACCATCACCGCCACAGATGCTGACACAGACACC---ACGGCCACCGGGATG 953
Qy 253 ValAlaTyrrSerIleHisSerGlnGluProLysAspProHisAspLeuMetPheThrIle 272
Db 954 GTGGGTACCGGATCGGTACCGACAGACCCACAGAGCCCGTCCCAAGATATGTTCAACCATC 1013
Qy 273 HisArgSerThrGlyThrIleSerValIleSerGlyLeuAspArgGluLysValPro 292
Db 1014 AACAGCGAGATCGAGATATCGTCACAGTGGCGGCTGCTGGGACCGAGAGAAAGTTCAG 1073
Qy 293 GluTyrrLeuThrIleGlnAlaThrAspMetAspGlyAsp-----GlySerThr 309
Db 1074 CAGTACACAGTCATCGTTACGGCCACAGATATGGAAGGAAATCTCAACTATGGCCTCTCA 1133
Qy 310 ThrThrAlaValAlaValGluIleLeuAspAlaAsnAspAsnAlaProMetPheAsp 329
Db 1134 AACACAGCCACGACCATCATCCGTCGACAGATGTGAATGACACCCCGTCAGAAATTACC 1193
Qy 330 ProGlnLysTyrrGluAlaHisValProGluAsnAlaValGlyHisGluValGlnArgLeu 349
Db 1194 GCCAGCAGTTTGCAGGGAGGTCCCGAAACAGCGTGGAGACCGGTGCGCAAACTC 1253
Qy 350 ThrValThrAspLeuAspAlaProAsnSerProAlaThrArgAlaThrTyrrLeuIleMet 369
Db 1254 ACGGTGATGGACCGAGATCAGCCCACTCTCCAAACTGGAATGCCGTATACCGCATCATC 1313
Qy 370 GlyGlyAspAspGlyAspHisPheThrIleThrHisProGluSerAsnGlnGlyIle 389
Db 1314 AGTGGGATCCATCCGGGCATTTAGCGTCGCGACAGCCCGTAAACCAAGAGGGCATG 1373
Qy 390 LeuThrThrArgLysGlyLeuAspPheGluAlaLysAsnGlnHisThrLeuTyrrValGlu 409
Db 1374 GTCACCGTGGTGAAGCGAGTCGACTACGAGCTCAACAGAGCTTTCATGCTGACAGTGTG 1433
Qy 410 ValThrAsnGluAlaProPheValLeuLysLeuPro-----ThrSerThrAlaThr 426
Db 1434 GTGTCAACACAGGCGCCCTCGCCAGCGGAATCCAGATGTCCTTCCAGTCCACGCGAGG 1493
Qy 427 IleValValHisValGluAspValAsnGluAlaProValPheValProProSerLysVal 446
Db 1494 GTGACCATCTCCATCATCATGACATCAACGAGGCTCCCTACTTCCCTCAAAACCAAGCTG 1553
Qy 447 ValGluValGlnGluGlyIleProThrGlyGluProValCysValTyrrThrAlaGluAsp 466
Db 1554 ATCCGCTCGAGGAGGGCGTGTCCCCCGCAGCCCGTGTGACCAAGTTCAGCTGTGGAC 1613
Qy 467 ProAspLys---GluAsnGlnLysIleSerTyrrArgIleLeuArgAspProAlaGlyTrp 485
Db 1614 CTTGACCGGTTCATGACAGCGGTGTGAGATACTCAAGCTGTGACAGCCAGCGAGTGG 1673
Qy 486 LeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThrLeuAspArgGluAsp 505
Db 1674 CTGCACATCAATGCCCAACGGCAGATCACACGGTGGCAGTGTGGACCGTGAG--- 1730
Qy 506 GluGlnPheValArgAsnAsnIleTyrrGluValMetValLeuAlaMetAspAsnGlySer 525
Db 1731 TCCCTCTCAACCAAAACACGTCTACAGGCCACCTTCCTGGCAGCTGACATGGGATA 1790
Qy 526 ProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAspValAsnAspHisGly 545
Db 1791 CCCCCGCGCAGCGGCACCGGACCCCTCCAGATCTATCTATTGACATCAACGACCAACGCC 1850
Qy 546 ProValProGluProArgGlnIleThrIleCysAsnGlnSerProValArgHisValLeu 565
Db 1851 CTTGAGCTGCTCCCAAGGAGCGCGCATCTGC---GAGAGGCCCAACCTGAAGCCCAATC 1907
Qy 566 AsnIleThr-----AspLysAspLeuSerProHisThrSerProPheGlnAlaGlnLeu 583
Db 1908 AACATCAGCGCGCGCGACGCTGACGTGCACCCCAACATCGGCCCTCTAGTCTTCGAGCTG 1967
Qy 584 ThrAspAspSerAspIleTyrrThrAlaGluValAenGlu----- 597
Db 1968 CCC-----TTTGTCCCGCGCGCGCTGCGGAAGAACTGGACCATCACCGCG 2012
Qy 598 ---GluGlyAspThrValValLeuSerLeuLys---LysPheLeuLysGlnAspThrTyrr 615
Db 2013 CTGAACGGTGTATGATGCCCACTAGCTTGGCATCTGTACTCTGGAGCGCGGATGTAT 2072
Qy 616 AspValHisLeuSerLeuSerAspHisGlyAsnLys-----GluGlnLeuThrValIle 633
Db 2073 GACGTCCTCCATCATCGTCACAGACTCTCGAAACCTCCCTCCCTGTCCACACATCGTCATC 2132
Qy 634 ArgAlaThrValCysAspCys-----HisGlyHisValGluThrCysProGlyProTrp 651
Db 2133 AAAGTCAGGTGTGCCATGTATGACAAACGGGAGTACTCACCATCTGGCGGAGTGCA 2192
Qy 652 LysGlyGlyPheIleLeuProValLeuGlyAlaValLeuAlaLeuLeuPheLeuLeu 671
Db 2193 GCGGCTGGTCTG-----GGCACCGTGCATCGTGGCATCTCTCATCTGCATCTC 2243
Qy 672 ValLeuLeuLeuVal----- 681
Db 2244 ATCTCTGTGACCATGTCTGTCTGTTCATGTGATGAAGCGCGGAGAGAGAGCGC 2303
Qy 682 LysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyrr 701
Db 2304 CACACGAGCAGCTGCTCATTCACCCCGAGCAGCGTCCGCGAAAGATCTCTCAAGTAT 2363
Qy 702 GlyGluGluGlyGlyGlyGluGluAspGlnAspTyrrAspIleThrGlnLeuHisArgGly 721
Db 2364 GACGAGGAGGGGTGGCGAGGAGGACGAGGACTACGACCTCAGCCATCGACCATGTGACGAGCG 2423
Qy 722 LeuGluAlaArgProGluVal-----ValLeuArgAsnAspValAla 735
Db 2424 ---GAAGCCATCGGCGACGTGCCAAGCAAGCCCTCGCGTGGTGGTGTGATGAGCGG 2480
Qy 736 ProThrIleProThrProMetTyrr-----ArgProArgProAlaAsnProAspGlu 753
Db 2481 CCG---GTGGGCGCTGAGCCCGCAGTACCGCATCGCGCCATGGTCCCGCAGCGCAG 2537
Qy 754 IleGlyAsnPheIleLeuLeuLysAlaAlaAsnThrAspProThrAlaProPro 773
Db 2538 ATCGGTGATCTCATCAATGAGGACTCCGCGCTGTGTACACAGCACCACCGCACCCCC 2597
Qy 774 TyrrAspThrLeuLeuValPheAspTyrrGluGlySerGlySerAspAlaAlaSerLeuSer 793
Db 2598 TATGACTCCCTGCTGTCTTCGACTACGAGGGGAGCGGCTCCACCGCAGGCTCCGCTCAGC 2657
```


QY 794 SerLeuThrSerSerAlaSerAspGlnAspGlnAspTyrAspTyrLeuAsnGluTyrGly 813
|||||
Db 2658 TCCCTGAACTCATCCAGTTCGGG---GACCAAGACTACGATTACCTCAAGACTGGGC 2714
|||||
QY 814 SerArgPheLysLeuAlaAspMetTyrGlyGlyGluAsp 829
|||||
Db 2715 CCCAGATTCAAGAGCTGGCGGACATGATGGAGGTGGTGAAGAGGAT 2762
|||||

RESULT 9
US-08-188-228-61
; Sequence 61 Application US/08188228
; Patent No. 5597725
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Borun
; STREET: 6300 Sears Tower, 233 S. Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/188,228
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/049,460
; FILING DATE: 19 APR 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,643
; FILING DATE: 17 APR 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5597725and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31340
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2690 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-08-188-228-61
Alignment Scores:
Pred. No.: 1 69e-95 Length: 2690
Score: 1025.50 Matches: 266
Percent Similarity: 51.89% Conservative: 119
Best Local Similarity: 35.85% Mismatches: 284
Query Match: 23.47% Indels: 73
DB: 1 Gaps: 22
US-09-916-849A-1 (1-829) x US-08-188-228-61 (1-2690)
QY 4 ProArgGlyProLeuAla---SerLeuLeuLeuGlnValCysTrpLeuGlnCysAla 22
|||||
Db 451 CCGAGACTCCGCTCGTCTGCTGCTCTCCAGGTGCTGCTGCTTAACATCTGCA 510
|||||
QY 23 AlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAlaGlyGly 42
|||||
|||||

Db 511 GAAGATTTGGACTGCTGCTGATTTTCAGCAGAAAAGTGTTCATATCAATCAGCCAGCT 570
QY 43 AlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGlyGlnGlu 62
|||||
Db 571 GAATTCATTGAGCACAGTCAATTCCTAACTTGACCTTCAGTCACTGTAGGAAACAGAC 630
|||||
QY 63 ProAlaLeuPheSerThrAspAsnAspAspPheThrVal--- 75
|||||
Db 631 AAGCTACGCTATGAGGTCTCGAGCCCATACTTCAAGGTCAACAGCGATGGGGCTTAGTT 690
|||||
QY 76 -----ArgAsnGlyGluThrValGlnGluArgArgSerLeuLysGluArgAsnPro--- 92
|||||
Db 691 GCTCTGAGAAACATAACTGCGAGTGGGCAAACTCTGTTGCTCCATGACAGGCCCCCAT 750
|||||
QY 93 -----LeuLysIlePheProSerLysArgIle----- 101
|||||
Db 751 GCGGAAGATATGCGCAGAACTCTGTATTCGCGGGGAAAAGACATCCAGGGTCTCTTCGAG 810
|||||
QY 102 -----LeuArgArgHisLysArgHisLysArgAspTrpValVal 111
|||||
Db 811 GATATATTAAATTTGCAAGAACTTCTCTGTCCCAAGACAAAAGAGGTCCATTTGGTGA 870
|||||
QY 112 AlaProIleSerValProGluAsnGlyLysGlyProPheProGlnArgLeuAsnGlnLeu 131
|||||
Db 871 TCTCCCATTTAAATTCAGAGATCAGAGACAGCCCTTTCCCAAGAGATGTTGGCAAGTA 930
|||||
QY 132 LysSerAsnLysAspArgAspThrLysIlePheTyrSerIleThrGlyProGlyAlaAsp 151
|||||
Db 931 ---GTCGATAGTACAGAGCCAGAAAGTCCAGTTCGCGCTCACTGGAAGAGTGGAT 987
|||||
QY 152 SerProGluGlyValPheAlaValGluLysGluThrGlyTrpLeuLeuAsnLys 171
|||||
Db 988 CAAGAGCCTTAAGGAATTTTCAATCAATGAGACACAGGAGGCGTCTCCGTGACAGCG 1047
|||||
QY 172 ProLeuAspArgGluGluIleAlaLysTyrGluLeuPheGlyHisAlaValSerGluAsn 191
|||||
Db 1048 ACCTTGGACAGAGAAGTAATCGCTCTTATCAACTATTGTTGGAGACCACTGATGTCAT 1107
|||||
QY 192 GlyAlaSerValGluAspProMetAsnIleSerIleValThrAspGlnAsnAspHis 211
|||||
Db 1108 GCGAAACTCTCGAGGGCGCGTCTCTGGAAGTCATTGTGATGATGATGATGATGATGAT 1167
|||||
QY 212 LysProLysPheThrGlnAspThrPheArgGlySerValLeuGluGlyValLeuProGly 231
|||||
Db 1168 CGACCGATCTTCGGGAAGGCCCTTACATCGGCCACGTCATGGAAGGTCACCCACAGGC 1227
|||||
QY 232 ThrSerValMetGlnValThrAlaThrAspGluAspAlaIleTyrThrTyrAsnGly 251
|||||
Db 1228 ACCACAGTATGCGGATGACAGCCCTTTGATGATGATGATGATGATGATGATGATGAT 1284
|||||
QY 252 ValValAlaTyrSerIleHisSerGlnGluProLysAspProHisAspLeuMetPheThr 271
|||||
Db 1285 CTCCTGCGGTATATATTCGTCACAGAGCGCTGCAAGCCATCTCCCAACATGTTCTAC 1344
|||||
QY 272 IleHisArgSerThrGlyThrIle---SerValIleSerSerGly---LeuAspArgGlu 289
|||||
Db 1345 ATCGATCTCGAAGAGGAGACATTTGCTACTGTTGTTCACCTCGCTGCTGCGACGAGAG 1404
|||||
QY 290 LysVal-----ProGluTyrThrLeuThrIleGlnAlaThrAspMetAspGly----- 305
|||||
Db 1405 ACTCTGGAAATCCCAAGTATGAACTGATCATCGAGGCTCAAGATATGCTGGATGGAT 1464
|||||
QY 306 AspGlySerThrThrAlaValAlaValGluIleLeuAspAlaAsnAspAsnAla 325
|||||
Db 1465 GTTGGATTAAACAGGCACGCCACAGCCACGATCATGATCGATGACAAAAATGATCACTCA 1524
|||||
QY 326 ProMetPheAspProGlnLysTyrGluAlaHisValProGluAsnAlaValGlyHisGlu 345
|||||
Db 1525 CCAAAATTCACCAAGAAAGTTCACGCCACAGTCGAGGAGGAGGCTGTGGGA---GTT 1581
|||||
QY 346 ValGlnArgGluThrValThrAspLeuAspAlaProAsnSerProAlaTrpArgAlaThr 365
|||||
Db 1582 ATTGTCAATTTGACAGTTGAAGATAAGGTGACCCACCACAGGTGCTATGAGGGCTGCC 1641
|||||

Qy 463 ThrAlaGluAspProAspLys---GluAsnGlnLysIleSerTyraGileLeuArgAsp 481
 Db 1942 AATGCCAGCGACCCGAGTCCCTGAGCATCAACACCATCAGTATTCCTTACAGGAC 2001
 Qy 482 ProAlaGlyThrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThrLeu 501
 Db 2002 CCAGCAGGTGCGTCAATTAATCAACCCATCAACCCGAGTGTGTGATCAACCCGAGGAGCTG 2061
 Qy 502 AspArgGluAspGluGlnPheValArgAsnAsnIleTyraGluValMetValLeuAlaMet 521
 Db 2062 GACCGTGAG---TCCCATTTGTGACACACAGCGGTACACTGCTCTCTTCTTGGCATT 2118
 Qy 522 AspAsnGlySerProProThrThrGlyThrLeuLeuLeuLeuLeuLeuLeuLeuLeuVal 541
 Db 2119 GACAGTGGCAACCCCTCCGCTACGCGACCTGGGACTTTGCTGATAACCCCTGGAGAGCTG 2178
 Qy 542 AsnAspHisGlyProValProGluProArgGlnIleThrLeuCys-----Asn 557
 Db 2179 AATGACAAAGCCCGTTCATTATCCACACAGCTGAGAGTCTGTGATGATGCCAAAGAAC 2238
 Qy 558 GlnSerProValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSer 577
 Db 2239 CTCAGTGTAGTC-----ATTTGGGAGCATCAGATTAAGATCTTCACCCGATACAGAT 2292
 Qy 578 ProPheGlnAlaGlnLeuThrAspAspSer-----AspIleTyraThrAlaGluVal 595
 Db 2293 CCTTCAAAATTTGAATCCCAACCAAGCTGTTCTGTATAAGTCTGG-----AAGATC 2346
 Qy 596 AsnGluGluGlyAspThr---ValValLeuSerLeuLysLysPheLeuLysGlnAspThr 614
 Db 2347 TCCAGATCAACAATACACACGCCCTGTGAGCTTCTTCAAAATCTGAACAAAGCAAC 2406
 Qy 615 TyrAspValHisLeuSerLeuSerAspHisGlyAsn-----LysGluGlnLeuThrVal 632
 Db 2407 TACAACCTCCCATCATGTGTGACAGATTTCAGGAAACCCACCCATGACGATATACAGAT 2466
 Qy 633 IleArgAlaThrValCysAspCysHisGlyHisValGluThrCysProGlyProTriLys 652
 Db 2467 CTCAGGTACAGTGTCTCTCTGAGGAATTCAGAGTGGNCTGCAACGGCGGGGGGCC 2526
 Qy 653 GlyGlyPheLeuProValLeuGlyValAlaValLeuAlaLeuLeuPheLeuLeuVal 672
 Db 2527 CTGGCTTCAGCTGCC-----TCAGTCTGCTCTCTGAGCTCTTCAGCTTAGCTGT 2580
 Qy 673 LeuLeuLeuLeuVal-----ArgLysLys 680
 Db 2581 CTGTGAGACCTCTGAGCTGTGAAGCTTGACTCCCAAGTTTCCATAGCAACAGGAAAAA 2640
 Qy 681 ArgLys 682
 Db 2641 AAAAAA 2646

RESULT 12
 US-08-474-067-3
 ; Sequence 3, Application US/08474067
 ; Patent No. 5811518
 ; GENERAL INFORMATION:
 ; APPLICANT: Ranscht, Barbara
 ; TITLE OF INVENTION: T-Cadherin Adhesion Molecule
 ; NUMBER OF SEQUENCES: 9
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Campbell and Flores
 ; STREET: 4370 La Jolla Village Drive, Suite 700
 ; CITY: San Diego
 ; STATE: California
 ; COUNTRY: United States
 ; ZIP: 92122
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/474,067
 FILING DATE: 07-JUN-1995
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/213,361
 FILING DATE: 14-MAY-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/607,293
 FILING DATE: 30-OCT-1990
 ATTORNEY/AGENT INFORMATION:
 NAME: Campbell, Cathryn A.
 REGISTRATION NUMBER: 31,815
 REFERENCE/DOCKET NUMBER: P-LJ 1682
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 535-9001
 TELEFAX: (619) 535-8949
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2779 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 40..2191
 US-08-474-067-3

Alignment Scores:
 Pred. No.: 1,6e-90 Length: 2779
 Score: 977.50 Matches: 265
 Percent Similarity: 49.81% Conservative: 127
 Best Local Similarity: 33.67% Mismatches: 309
 Query Match: 22.37% Indels: 86
 DB: 1 Gaps: 22

US-09-916-849A-1 (1-829) x US-08-474-067-3 (1-2779)

Qy 10 SerLeuLeuLeuGlnValCysTrpLeuGlnCysAlaAlaSerCluProCysArgAla 29
 Db 67 TCTTCTCTCTGTCAGAGTCTGCTGTCGTCAGAGATTTAGATGACCCCT 126
 Qy 30 ValPheArgGluAlaGluValThrLeuGluAlaGlyGlyAlaGluGlnGluProGlyGln 49
 Db 127 GGATTCACCAAAAGGTTTTTATATTAATGAACAGCCATTTGAATTCACAGACCCAGCCA 186
 Qy 50 AlaLeuGlyLysValPheMetGlyCysProGlyGlnGluProAlaLeuPheSerThrAsp 69
 Db 187 ATTCTGAACCTGCTGTTGATGACTGCAAGGGGAAATAACAAATTGAACCTTCAAGTTCT 246
 Qy 70 AsnAspAspPheThrValArgAsnGlyGluThrValGlnGluArgArgSerLeuLysGlu 89
 Db 247 AACCCAGACTTTAAGGTGGAACACGATGATCTTAGTTGCACTGAGATGATGATCAGAA 306
 Qy 89 ----- 89
 Db 307 GCTGGCAGAGCTTTGTTGTCATGCAGCGTCTGAGCATGCTGAGGATATGCGAGAAAT 366
 Qy 90 -----ArgAsnProLeuLys---IlePheProSerLysArg 100
 Db 367 TTGATTGTTGGAGCTGATGAGAGCAGCAGCATGATTAAGGAAATCTTTAAGATAGAGGC 426
 Qy 101 IleLeu-----ArgArgHisLysArgAspTrpValValAlaProIleSerValProGlu 118
 Db 427 AACCTTGAATTCACAGACAAAAAGGGCTATCTCTGGGACTCCAATAATTAATTCAGAA 486
 Qy 119 AsnGlyLysGlyProPheProGlnArgLeuAsnGln---LeuLysSerAsnLysAspArg 137
 Db 487 AATCAAGACCCACCATTTCCAGATCAGTTCGCAAGGTCATCAGGAGTGAAGGACAGAG 546
 Qy 138 AspThrLysIlePheTyraSerIleThrGlyProGlyValAlaAspSerProProGlyVal 157
 Db 547 GGAGCAAG-----TTCCGACTCTCTGTAAGGGAGTAGATCAAGACCCCGAAGGAAT 600

QY 158 PheAlaValGluLysGluThrClyTrpLeuLeuLeuAsnLysProLeuAspArgGluGlu 177
 DB 601 TTTAGAAATCAATAGATCAGTGGGATGTCTGTGACCCGAGCCCTGGATAGAGAAGCA 660
 QY 178 IleAlaLysTyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAsp 197
 DB 661 ATAGCCATTATGAGTGAAGTGAAGTAAACCGATTAAAGTGGAATAATCATGATGGC 720
 QY 198 ProMetAsnIleSerIleValThrAspGlnAsnAspHisLysProLysPheThrGln 217
 DB 721 CCAGCCCGCCAGATATTTCTTTATTGATCAAAATGATAACAGCCGATGTTCAAAAGA 780
 QY 218 AspThrPheAspGlySerValLeuGluValLeuProGlyThrSerValMetGlnVal 237
 DB 781 GGACCTATGTTGTCACGTCATGGAAGATCCCTACAGGAACAACCTGTGATGGCGATG 840
 QY 238 ThrAlaThrAspGluAspAlaIleTyrThrTyrAsnGlyValAlaIleTyrSerIle 257
 DB 841 ACAGCATTTGATGCTGATGATCCT---AGCAGACACACGCTCTCTCGCGTATAACATC 897
 QY 258 HisSerGlnGluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGly 277
 DB 898 CTCAGACAGACACCTTACCAACCTTCCCAATATGTTCTACATGTGACCCAGAAAAGGGA 957
 QY 278 ThrIleSerValIleSerSerGly-----LeuAspArgGlu-----LysValProGlu 293
 DB 958 GATATTGTACAGTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1017
 QY 294 TyrThrLeuThrIleGlnAlaThrAspMetAspGlyAsp-----GlySerThrThr 311
 DB 1018 TACGAGCTGGTATTGAAGCAAGGATATGGCGCGCATGATGGGACCTTACTGGAACT 1077
 QY 312 AlaValAlaValGluIleLeuAspAlaAsnAspAsnAlaProMetPheAspProGln 331
 DB 1078 GCACTGCGCATTTCTTTATTGATGACAAACAGCACCACCCAGAAATTTACCAAGAAG 1137
 QY 332 LysTyrGluAlaHisValProGluAsnAlaValGlyHisGluValGlnArgLeuThrVal 351
 DB 1138 GAGTTTCCAGGCGCACAGTAAAGAAAGGAGTCAAGGA---GTAATAGTAAACTTTAACTGTT 1194
 QY 352 ThrAspLeuAspAlaProAsnSerProAlaThrArgAlaThrTyrLeuIleMetGlyGly 371
 DB 1195 GGTGACCGAGATGACCCAGCACTGGAGCAGTGGAGAGCTGTACACTATTATTAAACGGA 1254
 QY 372 AspAspGlyAspHisPheThrIleThrHisProGluSerAsnGlnGlyIleLeuThr 391
 DB 1255 AATCCAGGCGCAGAGTTTGAATATCCATACCAATCCCAAGCACTAATGAGGAATGCTCTCT 1314
 QY 392 ThrArgLysGlyLeuAspPheGluAlaLysAsnGlnHisThrLeuTyrValGluValThr 411
 DB 1315 GTTGTCAAAACCTTTAGATATGATGATTTTCAAGTTTCAAGTTTCAAGTTTCAAGTTT 1374
 QY 412 AsnGluAlaProPheValLeuLysLeu-----ProThrSerThrAlaThrIleVal 428
 DB 1375 AATGAAGACCCGTTGATTCCAGACATAGCTACGCTCCAGTTCCAGCAACAGATTCAG 1434
 QY 429 ValHisValGluAspValAsnGluAlaProValPheValProProSerLysValValGlu 448
 DB 1435 ATCACCCTTGAGATGTAATCAAGGCGCTGTTTCCACCACCAACCCCAATGACAGTGACA 1494
 QY 449 ValGlnGluGlyIleProThrGlyGluProValCysValTyrThrAlaGluAspProAsp 468
 DB 1495 AAACAGAGAACATCCCTATTGGCAGCATTTGTTTAAACAGTAATGCCACTGATCCAGAT 1554
 QY 469 Lys-----GluAsnGlnLysIleSerTyrArgIleLeuArgAspProAlaGlyTrpLeuAla 487
 DB 1555 ACTTTTGCAACATCAGCATATCAGGTATTTCAGTTTACAGGATCCAGCAAGCTGGCTAGAG 1614
 QY 488 MetAspProAspSerGlyGlnValThrAlaValGlyThrLeuAspArgGluAspGluGln 507
 DB 1615 ATTATCCCAACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1674
 QY 508 PheValArgAsnAsnIleTyrGluValMetValLeuAlaMetAspAsnGlySerProPro 527

DB 1675 ---GTTCCAGGATAACAAATACACTGCTCTCTCTCTGCAATAGACAGTGGTAACCTCTCT 1731
 QY 528 ThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAspValAsnAspHisGlyProVal 547
 DB 1732 GCTACAGGTACAGAACTTTTACATCACTTCTGGAGGACGTCATCAATCAATGATCCCTCC 1791
 QY 548 ProGluProArgGlnIleThrIleCysAsnGlnSerProValArgHis-----ValLeu 565
 DB 1792 CTTTACCACAACTGGCAAAAGTCTGTGATGATGATGATGATGATGATGATGATGATGAT 1851
 QY 566 AsnIleThrAspLysAspLysSerProHisThrSerProPheGlnAlaGlnLeuThrAsp 585
 DB 1852 CGAGCATCAGACAAAGACCTCCATCCCAACACAGATCCATTTAAATTTGAATGAGTAAAG 1911
 QY 586 AspSerAspIleTyrTrpThrAlaGluValAsnGluGlyAspThr---ValValLeu 604
 DB 1912 CAATCTGGTCCGAAAAGATTATGGAATACAAAGCTTAAACATCTCATGCTCCAGGTT 1971
 QY 605 SerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSerLysSerAspHis 624
 DB 1972 GTCTGCTTCAAAACCTGAAAAGGCCAATTACAAACATCCCAATCTCAGTGACAGATTCT 2031
 QY 625 GlyAsnLysGluGlnLeu-----ThrValIleArgAlaThrValCysAspCys-Hi 641
 DB 2032 GGA---AAACCACTCTGACTTAACACACAGCACTGAAATTTCAAGTGTGTTCTGCAAG 2088
 QY 641 scgIHisValGluThrCysProGlyProTyrIleGlyGlyPheIleLeuProValLeuGlu 661
 DB 2089 AAT-----CCAGATGGACTCGACTGCAAGTGAAGTGGTCTTCTTATGTT 2130
 QY 661 yAlaValLeu-----LeuValLeuLeuLeuValArgLysLysArgLysLysLysGluPr 686
 DB 2131 AGCATGACTCTTATCT 2190
 QY 671 -----LeuValLeuLeuLeuValArgLysLysArgLysLysLysLysLysGluPr 686
 DB 2191 TAAGCATTCAGCTTATTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2248
 QY 686 oLeuLeuLeuProGluAspThrArgAspAsnValPheTyrTyrGlyGluGlyGlu 706
 DB 2249 -----AGAAATTAACCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2283
 QY 706 yGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisArgGlyLeuGluAlaArgPr 726
 DB 2284 TATACAGAGTAGGAGGGAGGGAGGATTTTTCACAGTCAAAAATAGCAACAAATGCCGG 2341
 QY 726 oGluValValLeuArgAsn 732
 DB 2342 -GTTGTCAAATTAAGAAAT 2359

RESULT 13

US-08-474-068A-3
 ; Sequence 3, Application US/08474068A
 ; Patent No. 5837525
 ; GENERAL INFORMATION:
 ; APPLICANT: Ranscht, Barbara
 ; TITLE OF INVENTION: T-cadherin Adhesion Molecule
 ; NUMBER OF SEQUENCES: 9
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: Campbell & Flores LLP
 ; STREET: 4370 La Jolla Village Drive, Suite 700
 ; CITY: San Diego
 ; STATE: California
 ; COUNTRY: United States
 ; ZIP: 92122
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/474,068A

```

; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/213,361
; FILING DATE: 14-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/607,293
; FILING DATE: 30-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1683
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2779 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 40..2191
; US-08-474-068A-3

Alignment Scores:
Pred. NO.: 1,6e-90 Length: 2779
Score: 977.50 Matches: 265
Percent Similarity: 49.81% Conservative: 127
Best Local Similarity: 33.67% Mismatches: 309
Query Match: 22.37% Indels: 86
DB: 2 Gaps: 22

US-09-916-849A-1 (1-829) x US-08-474-068A-3 (1-2779)
QY 10 SerLeuLeuLeuGlnValCysTrpLeuGlnCysAlaLaserGluProCysArgAla 29
DB 67 TCCTTTCTGCTGCCAGGTTCTGCTTGGTGTGCAGAAATTAGAAATGACCCCT 126
QY 30 ValPheArgGluAlaGluValThrLeuGluAlaGlyAlaGluGlnProGlyGln 49
DB 127 GGATTCCAGCAAAAGGTTTTTATATTGAACAGCCATTGAATTCACAGAGGACGCCA 186
QY 50 AlaLeuGlyLysValPheMetGlyCysProGlyGlnGluProAlaLeuPheSerThrAsp 69
DB 187 ATTCTGAACCTGGTGTGGTACTGCAAGGGGAATAACAAATTGAACCTCGAAGTTCT 246
QY 70 AsnAspPheThrValArgAsnGlyGluThrValGlnGluArgArgSerLeuLysGlu 89
DB 247 AACCCAGACTTTAAGGTGGAAACAGCATGGATCTTTAGTTGCACCTGAAGATGTATCAGAA 306
QY 89 ----- 89
DB 307 GCTGGCAGAGCTTTGTTTGTCCATGCAACGCTCTGACCATCTGAGGATATGGCAGAAATT 366
QY 90 -----ArgAsnProLeuLys---IlePheProSerLysArg 100
DB 367 TTGATTCTGGAGCTGATGAGACGACGATGCATTAAAGGAATCTTTAAGATAGAAGGC 426
QY 101 IleLeu-----ArgArgHisLysArgAspTrpValValAlaProIleSerValProGlu 118
DB 427 AACCTTGAATTCACAGACAAAAGGGCTATTCTGGCGACTCCAATATTAATTCAGAA 486
QY 119 AsnGlyLysGlyProPheProGlnArgLeuAsnGln---LeuLysSerAsnLysAspArg 137
DB 487 AATCAAGACACCATTTCCAGATCAGTTGGCAAGGTGCATCAGGAGTGAAGGCAGAG 546
QY 138 AspThrLysIlePheTrpSerIleThrGlyProGlyAlaAspSerProGluGlyVal 157
DB 547 GGAGCAAG-----TTCCGACTCTCTGGTAAGGAGTAGATCAAGACCCGAAAGGAATT 600
QY 158 PheAlaValGluLysGluThrGlyTrpLeuLeuAsnLysProLeuAspArgGluGlu 177

```

```

DB 601 TTTAGAAATCAATGAGATCAGTGGGGATGCTCTGTGACCCGACCCCTGATAGAGAGCA 660
QY 178 IleAlaLysTrpGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAsp 197
DB 661 ATAGCCATATTAGCTGGAAGTTGAAGTAACGGGATTAAAGTGGGAAAATCATGTATGGC 720
QY 198 ProMetAsnIleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGln 217
DB 721 CCAGTCCGCGCTAGATATTCTGTTATTGATCAAAATGATAACAGCCGATGTCAAGAA 780
QY 218 AspThrPheArgLysSerValLeuGluGlyValLeuProGlyThrSerValMetGlnVal 237
DB 781 GGACCCCTATGTTGGTCACGTCATGGAAGGATCCCTTACAGGAACAACACTGTATGCGGATG 840
QY 238 ThrAlaThrAspGluAspAlaIleThrTrpAsnGlyValValAlaValSerIle 257
DB 841 ACAGCATTTGATGCTGATCTCT---ACGACAGACAGCCTCTTCTGCGGTATACATC 897
QY 258 HisSerGlnGluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGly 277
DB 898 CTCAGCAGACACCTACCAACCTTCCCAATATGTTCTACATTGACCCAGAAAGGGA 957
QY 278 ThrIleSerValIleSerSerGly-----LeuAspArgGlu-----LysValProGlu 293
DB 958 GATATTGTCCAGTGGTGTGCGCTGTACTGCTGGATCGTGAGACAATGGAAACGCCGAAG 1017
QY 294 TyrThrLeuThrIleGlnAlaThrAspMetAspGlyAsp-----GlySerThrThrThr 311
DB 1018 TACGAGCTGGTATTGAAGCCCAAGGATATGGCGGCCCATGATGTGGGACTTACTGGAAC 1077
QY 312 AlaValAlaValValGluIleLeuAspAlaAsnAspAsnAlaProMetPheAspProGln 331
DB 1078 GCAACTGCCACTATTCTTATTGATGACAAAAACGACCCACCAGAAATTTACCAAGAG 1137
QY 332 LysTrpGluAlaHisValProGluAsnAlaValGlyHisGluValGlnArgLeuThrVal 351
DB 1138 GAGTTTCAGCCACAGTAAGGAAGAGTCACAGGA---GTAATAGTAACCTTAACCTGTT 1194
QY 352 ThrAspLeuAlaProAsnSerProAlaTrpArgAlaThrTyrLeuIleMetGlyGly 371
DB 1195 GGTGACCGAGATGACCCAGCAACTGGAGCATGGAGAGCTGTCTACACTATTATTAACGGA 1254
QY 372 AspAspGlyAspHisPheThrIleThrHisProGluSerAsnGlnGlyIleLeuThr 391
DB 1255 AATCCAGGCGCAGAGTTTGAATCCATCAATCCAGACTCCAGACTAATGAGGGAATGCTCTCT 1314
QY 392 ThrArgLysGlyLeuAspPheGluAlaLysAsnGlnHisThrLeuTyrValGluValThr 411
DB 1315 GTTGTCAAACCTTTAGACTATGAGATTTTCAGCATTTTCACACTTCTGTGATAAAGTAGAA 1374
QY 412 AsnGluAlaProPheValLeuLysLeu-----ProThrSerThrAlaThrIleVal 428
DB 1375 AATGAAGACCCCTTGATTCCAGACATAGCCTACGCTCCAGTTCCACAGCAACAGTTCCAG 1434
QY 429 ValHisValGluAspValAsnGluAlaProValPheValProSerLysValValGlu 448
DB 1435 ATCACCCTTGAGATGTAAGAGCCCTGTTTCCACCACCAACCCCAATGACATGACAG 1494
QY 449 ValGlnGluGlyIleProThrGlyGluProValCysValTyrThrAlaGluAspProAsp 468
DB 1495 AAAACAAGAACATCCCTATTGGCAGCATTTGTTTAACAGTAAATGCCACTGATCCAGAT 1554
QY 469 Lys---GluAsnGlnLysIleSerTrpArgIleLeuArgAspProAlaGlyTrpLeuAla 487
DB 1555 ACTTTGCCAATCACTATCAGGTATTGATTCAAGGATCCAGCAAGTGGGTAGAG 1614
QY 488 MetAspProAspSerGlyGlnValThrAlaValGlyThrLeuAspArgGluAspGluGln 507
DB 1615 ATTAATCCCAACCAATGTTACCGTTGCCACCACCTGCTGCTGCTGGATCGGGAATCTCCG 1674
QY 508 PheValArgAsnAsnIleThrGluValMetValLeuAlaMetAspAsnGlySerProPro 527

```


FILING DATE: HEREWITH

CLASSIFICATION:

CLASSIFICATION:
PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Zeller, Karen J

REGISTRATION NUMBER: 37,071

REFERENCE/DOCKET NUMBER: PA-0002 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (650) 855-0555

TELEPHONE: (650) 845-4166
TELEFAX: (650) 845-4166

INFORMATION FOR SEO ID NO: 1125:

SEQUENCE CHARACTERISTICS:

LENGTH: 2880 base pairs

LENGTH: 2000 base pairs
TYPE: nucleic acid

LIFE: NUCLEIC ACID
STRANDEDNESS: SINGLE

STRANDEDNESS: single
TOPOLOGY: linear

TOPOLOGY: Linear
IMMEDIATE SOURCE:

IMMEDIATE SOURCE:
LIBRARY. GENBRANK

LIBRARY: GENBANK
CLONE: J1389852

CLONE: g134
016-A34-1135

ent Scores:

2.15e-90 Length: 2880

Matches:
976.50

t Similarity: 50.51%

Local similarity: 34.19%

Match: 22.35%

MacOS: 11.0.0
Gaps: 4

916-849A-1 (1-829) X US-09-016-434-1125 (1-2880)

```

90 ArgAsnProLeuTys---llePheProSerLysArgIleLeuArgHrgHisLysArgAsp 108
      |||
416 AGAAACCAACCAACACATTAAGGTGAACACCAAGTCCATCATGCTCCCAAAAGGGGA 475
      |||
109 TrpValValAlaProIleSerValProGluAsnGlyLysGlyProPheProGlnArgLeu 128
      |||
476 TGGGTATGGAAATCAGTTCTTTGTTTATAGAAAGACATATGGAGCACAGATCTCAGTATGTT 535
      |||
129 AsnGlnIleLysSerAsnLysAspArg--AspThrLysIlePheTyrSerIleThrGly 147
      |||
536 GGAAAGCTGCACTCCATTTCTGACAAAGGTGATGGATCTGTCAAGTACATCCTTACTGGA 595
      |||
148 ProGlyAlaAspSerProProGluGlyValPheAlaValGluLysGluThrGlyTrpLeu 167
      |||
596 GAGGTGCTGGGACT-----ATATTTATCATTCAGCATACCAACGGGTGATATC 643
      |||
168 LeuLeuAsnLysProLeuAspArgGluLuealalysTyrGluLeuPheGlyHisAla 187
      |||
644 CACTCAACAAAAGACCTTAGACAGAGACGACGAAGACCACCTATGTGTTTATGCTCAAGT 703
      |||
188 ValSerGluAsnGlyAlaSerValGluAspProMetAsnIleSerIlelle---ValThr 206
      |||
704 ATTGATAGACGTACAAACAACACCTTTGAGCCTGAATCCGAGTTTCATCAAGTGCAG 763
      |||
207 AspGlnAsnAspHisLysProLysPheThrGlnAspThrPheArgGlySerValLeuGlu 226
      |||
764 GACATCAATGACACCGCTCCAAAATTCACAGATGGACCATACATGTTATGTGCTCGTAA 823
      |||
227 GlyValLeuProGlyThrSerValMetGlnValThrAlaThrAspGluAspAlaIle 246
      |||
824 ATGTCAGATATGGGTACCTCTGTTCTACAGGTGACGCTTACTGATGCAGATGACCTACC 883
      |||
247 TyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGlnGluProLysAspProHis 266
      |||
884 TATGAAACACGGCTCGGTGGTTTACAGCATTTCTCCAGGGACAAACCC----- 931
      |||
267 AspLeuMetPheThrIleHisArgSerThrGlyThrIleSerValIleSerSerGlyLeu 286
      |||
932 -----TACTTCTCGCTCGACCTTCAAAACAGGAGTTATTAGACGGCTTTACATAACATG 985

```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 23, 2004, 00:01:04 ; Search time 933 Seconds
(without alignments)
4484.816 Million cell updates/sec

Title: US-09-916-849A-1
Perfect score: 4369
Sequence: 1 MGLPRGPLASLLQLQVCWLQ.....NEWGSRFKKLADMYGGGEDD 829

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 3327077 seqs, 2523723180 residues

Total number of hits satisfying chosen parameters: 6654154

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh

-Q=/cgn2_1/USPTO.spool.h/US09916849/runat_21092004_145806_7118/app.query.fasta_1.967

-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1

-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62

-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=ptc -THR MAX=100

-MAXLEN=2000000000 -USER=US09916849@cgn_11_593_@runat_21092004_145806_7118

-NCFU=6 -ICFU=3 -NO_WMAP -LARGESCORE -NEG SCORES=0 -WAIT -DSPELOCK=100

-LONGLOG -DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5

-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*

2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*

3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*

4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*

5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*

6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*

7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*

8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*

9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*

10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*

11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*

12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*

13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*

14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*

15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*

16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*

17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*

18: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*

19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

ALIGNMENTS

RESULT 1
US-09-975-723A-2
; Sequence 2, Application US/09975723A
; Publication No. US20030108529A1
; GENERAL INFORMATION:
; APPLICANT: Nackman, Gary
; APPLICANT: Foty, Ramsey
; TITLE OF INVENTION: Improvement of Endothelial Cell-Cell
; FILE OF INVENTION: Cohesion
; FILE REFERENCE: 601-1-101N
; CURRENT APPLICATION NUMBER: US/09/975,723A
; PRIOR FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: 60/241,216
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/243,693
; PRIOR FILING DATE: 2000-10-27
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 3170
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-975-723A-2

1	4369	100.0	3170	10	US-09-975-723A-2	Sequence 2, Appl.i
2	4369	100.0	3171	9	US-09-905-983-49	Sequence 49, Appl
3	4369	100.0	3171	9	US-09-880-107-3759	Sequence 3759, Ap
4	4369	100.0	3171	13	US-10-058-270A-125	Sequence 125, App
5	4369	100.0	3171	13	US-10-343-887-669	Sequence 669, App
6	4369	100.0	3171	13	US-10-172-118-669	Sequence 669, App
7	4369	100.0	3171	15	US-10-174-677-115	Sequence 115, App
8	4369	100.0	3171	15	US-10-158-123-2	Sequence 2, Appli
9	4369	100.0	3171	16	US-10-295-027-782	Sequence 782, App
10	4369	100.0	3171	16	US-10-295-027-799	Sequence 799, App
11	4369	100.0	3171	16	US-10-295-027-862	Sequence 862, App
12	4369	100.0	3171	16	US-10-295-027-895	Sequence 895, App
13	4369	100.0	3171	16	US-10-305-720-1348	Sequence 1348, Ap
14	4369	100.0	3171	17	US-10-775-920-188	Sequence 188, App
15	4369	100.0	3177	13	US-10-116-802-6	Sequence 6, Appli
16	4361	99.8	3205	13	US-10-229-345-7	Sequence 7, Appli
17	4361	99.8	3205	13	US-10-274-177-7	Sequence 7, Appli
18	4361	99.8	3205	15	US-10-301-822-17	Sequence 17, Appl
19	4361	99.8	3205	16	US-10-295-027-337	Sequence 337, App
20	4361	99.8	3205	16	US-10-295-027-1029	Sequence 1029, Ap
21	4361	99.8	3205	16	US-10-087-080-13	Sequence 13, Appli
22	4361	99.8	3205	17	US-10-650-112-7	Sequence 7, Appli
23	4361	99.8	3205	17	US-10-775-920-186	Sequence 186, App
24	4348	99.5	3219	9	US-09-895-652-17	Sequence 17, Appl
25	4310	98.6	3632	17	US-10-775-920-187	Sequence 187, App
26	3660	83.8	3314	16	US-10-062-674-1863	Sequence 1863, Ap
27	2420	55.4	4778	13	US-10-342-887-1027	Sequence 1027, Ap
28	2420	55.4	4778	13	US-10-172-118-1027	Sequence 1027, Ap
29	2420	55.4	4778	15	US-10-174-677-116	Sequence 116, App
30	2420	55.4	4828	15	US-10-097-340-36	Sequence 36, Appl
31	2420	55.4	4828	15	US-10-007-926A-328	Sequence 328, App
32	2420	55.4	4828	16	US-10-439-388-38	Sequence 38, Appl
33	2420	55.4	4895	15	US-10-198-846-11387	Sequence 11387, A
34	2409	55.1	4793	13	US-10-435-114-26834	Sequence 26834, A
35	2398	54.7	2768	9	US-09-905-983-4	Sequence 4, Appli
36	2388	54.7	2768	9	US-09-905-983-6	Sequence 6, Appli
37	2388	54.7	4396	12	US-10-152-319A-1900	Sequence 1900, Ap
38	2349.5	53.8	2808	14	US-10-165-049-1	Sequence 1, Appli
39	2342.5	53.6	2808	9	US-09-905-983-47	Sequence 47, Appl
40	2342.5	53.6	2808	16	US-10-173-551-33	Sequence 33, Appl
41	1714	39.2	3266	10	US-09-814-353-20361	Sequence 20361, A
42	1713	39.2	3427	14	US-10-044-090-589	Sequence 589, App
43	1713	39.2	3427	15	US-10-208-408-11	Sequence 11, Appl
44	1713	39.2	4122	16	US-10-133-937-31	Sequence 31, Appl
45	1713	39.2	4122	16	US-10-159-563-31	Sequence 31, Appl

421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
 Db CCAACCTCCACAGCCACCAATAGTGTCCAGTGGAGATGTGATGAGGACCTGTGT 1373
 441 ValProSerSerValValGluValGluGluIleProThrGluProValCys 460
 Db GTCCACCCCTCCAAAGTCGTTAGGTCAGAGGGGATCCCCCTCGGGAGCCCTGTGT 1433
 461 ValTyrThrAlaGluAspProAspLysGluAsnLysIleSerTyrArgIleLeuArg 480
 Db GTCTACACTGCAGAGAGCCCTGCAAGAGAGATCAAAAGATCACTACCCATCTTGAGA 1493
 481 AspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
 Db GACCCAGCAGGTCGTAGCATGACCCAGACAGTGGCAGGTCAAGCTGTGGGCACC 1553
 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
 Db CTCGACCGTGAAGTATGAGCATTTGTGAGGAACACATCTATGAAGTCATGGTCTTGCC 1613
 521 MetAspAsnGlySerProThrThrGlyThrLeuLeuLeuThrLeuIleAsp 540
 Db ATGACATATGAGAGCCCTCCACACCTGGCAGGAAACCTCTGCTAACACTGATGAT 1673
 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
 Db GTCAACGACCATGTCCTGAGCCCGTCAGATCACTCTGCAACCAAGAGCCCT 1733
 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
 Db GTGGCCACAGTGTGAACATCACGACAGGACCTGTCTCCACACCTCCCTTCCAG 1793
 581 AlaGlnLeuThrAspAspSerAspIleTyrThrAlaGluValAsnGluGlyAsp 600
 Db GCCCAGCTCACAGATGACTCAGACATCTACTGGAGCGCAGAGGTCAACAGGAGGTGAC 1853
 601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
 Db ACAGTGTCTTGTCTCCCTGAAGAGTCTCTGAAGCAGGATCATATGACGTGCACCTTCT 1913
 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
 Db CTGTCTGACCATGCAACAAAGAGCAGTGCAGGTGATGACGGCCACTGTGTGGAGCTGC 1973
 641 HisGlyHisValGluThrCysProGlyProTyrLysGlyGlyPheIleLeuProValLeu 660
 Db CATGCCCATGTGCAAAACCTGACCTGGACCTGGAAAGAGGTTTCATCTCTCCCTGTGCTG 2033
 661 GlyAlaValLeuAlaLeuPheLeuLeuValLeuLeuLeuValArgLysLys 680
 Db GGCGTGTCTCTGCT 2093
 681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
 Db CGGAAGATCAAGAGCCCTCTACTCCCAAGATGACACCCGTGACCAAGCTCTCTCTAC 2153
 701 TyrGlyGluGluGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
 Db TATGCGAAGAGGGGGTGGGAGAGGACCAAGGACTATGACATCAACCCAGCTCCACGA 2213
 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
 Db GTCTGTGAGGCGCAGCCGAGGTGTTCTCCCAATGACGTGGCAGCAACCAACCATCCCG 2273
 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
 Db ACACCATGTACCGTCTAGGCGACCAACCCAGATGAAATCGGCACTTTATTAATTGAG 2333
 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
 Db AACCTGAAGCGGCTTAACACAGACCCACAGCCCGCCCTACGACACCCCTCTGTGTGTC 2393

781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
 Db GACTATGAGGCGAGCGGCTCCGAGCGCGCTCCCTGAGCTCCCTCAGCTCCCTCGCTCC 2453
 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluIleProGlySerArgPheLysLysLeuAla 820
 Db GACCAAGACCAAGATGATGATATCTGACAGTGGGCGAGCCCTTCAAGAACCTGGCA 2513
 821 AspMetTyrGlyGlyGlyGluAspAsp 829
 Db GACATGTACGCTGGCGGAGGAGGAC 2540
 RESULT 3
 US-09-880-107-3759
 ; Sequence 3759, Application US/09880107
 ; Patent No. US20020142981A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Horne, Darci T.
 ; APPLICANT: Vockley, Joseph G.
 ; APPLICANT: Scherf, Uwe
 ; APPLICANT: Gene Logic, Inc.
 ; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
 ; FILE REFERENCE: 44921-5028-WO
 ; CURRENT APPLICATION NUMBER: US/09/880,107
 ; CURRENT FILING DATE: 2001-06-14
 ; PRIOR APPLICATION NUMBER: US 60/211,379
 ; PRIOR FILING DATE: 2000-06-14
 ; PRIOR APPLICATION NUMBER: US 60/237,054
 ; PRIOR FILING DATE: 2000-10-02
 ; NUMBER OF SEQ ID NOS: 3950
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 3759
 ; LENGTH: 3171
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: Genbank Accession No. US20020142981A1 X63629
 US-09-880-107-3759
 Alignment Scores:
 Pred. No.: 0 Length: 3171
 Score: 4369.00 Matches: 829
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 9 Gaps: 0
 US-09-916-849A-1 (1-829) x US-09-880-107-3759 (1-3171)
 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuGlnValCysTrpLeuGln 20
 Db 54 ATGGGCTCCCTCGTGGACCTCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 113
 21 CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
 Db 114 TGGCGGCTCCGAGCCGCTCGCGGGGTCTTCAGGAGGCTGAAGTACCTTTGGAGCG 173
 41 GlyGlyValGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
 Db 174 GGAGGCGGAGCAGGAGCCCGCCAGGCGCTGGGGAAGTATTCATGGCTGCCCTGGG 233
 61 GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr 80
 Db 234 CAAGACCCAGCTCTGTTAGCACTGATATGATGATGATGATGATGATGATGATGATGATG 293
 81 ValGlnGluArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
 Db 294 GTCCAGAAAGAGGTCTCACTGAAGGAAGAAATCCATTCAGATCTCCCATCCAAAGCT 353
 101 IleLeuArgHisLysArgAspTyrValValAlaProIleSerValProGluAsnGly 120
 Db 354 ATCTTACGAAGACACAAAGAGATTTGGTGGTGTCTCCATATCTCTCTCTCTCTCTCTCT 413

QY 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
DB 414 AAGGGTCCCTTCCCCAGAGACTGAATCAGCTCAAGTCTAATAAAGATAGAGACCAAG 473
QY 141 IlePheTyrSerIleThrGlyProGlyValAlaAspSerProGluGluValPheAlaVal 160
DB 474 ATTTTCTACAGACTACCGGGCGGGGCGAGACAGCCCCCTGAGGGTGTCTTCGCTGTA 533
QY 161 GluLysGluThrGlyTyrTrpLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys 180
DB 534 GAGAAAGAGAGACAGCTGGTGTGTTGAATAAGCACTGACCGGAGGAGATTTGCCAAG 593
QY 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
DB 594 TATGAGCTCTTTGGCCAGCTGTGTACAGAAATGGTGCTCAGTGGAGGACCCCATGAAC 653
QY 201 IleserIleleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
DB 654 ATCTCCATCATCGTACCGGACCAAGATGACCAGGCCCAAGTTTACCCAGGACACCTTC 713
QY 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
DB 714 CGAGGGAGTGTCTTAGAGGAGTCTTACCAGGTACTTCTGTGATGCAGGTGACAGCCACA 773
QY 241 AspGluAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln 260
DB 774 GATGAGGATGATGCACTACACCTACCAATGGGGTGTGTCTTACTCCATCCATAGCCAA 833
QY 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
DB 834 GAACAAAGACCCACACAGACCTCATGTTCACAATTCACCGAGCACAGGACCATCAGC 893
QY 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
DB 894 GTCATCTCCAGTGGCTGGACCGGGAAAAAGTCCCTGAGTACACACTGACCATCCAGGCC 953
QY 301 ThrAspMetAspGlyAspGlySerThrThrThrAlaValAlaValValGluIleLeuAsp 320
DB 954 ACAGACATGATGGGAGCGCTCCACCACCGCAGTGGCAGTAGTGGAGATCCTTGAT 1013
QY 321 AlaAsnAspAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
DB 1014 GCCAATGACATGCTCCCATGTTGACCCCGAGNAGTACAGGCCCATGTGCTTGAGAT 1073
QY 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
DB 1074 GCAGTGGGCCATGAGGTGCAGAGGCTCAGCGTCACTGATCTGACGCGCCCACTCACCA 1133
QY 361 AlaTrpArgAlaThrTyrLeuIleMetClyGlyAspAspGlyAspHisPheThrIleThr 380
DB 1134 GCGTGGCGTGCCACTACCTTATCANGGGCGGTGACACGGGACCATTTTACATCACC 1193
QY 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
DB 1194 ACCCACCTTGAGAGCAACAGGGCATCTTGACAACAGAGAGGGTTTGGATTTGAGGCC 1253
QY 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
DB 1254 AAAAAACAGCACACCTGTGCTGAGTTGAAGTGAACCAACAGAGCCCTTTTGTGCTGAAGCTC 1313
QY 421 ProThrSerThrAlaThrIleValIleValHisValGluAspValAsnGluAlaProValPhe 440
DB 1314 CCAACCTCCACAGCACCATAGTGTGTCCAGTGGAGGATGTGAATGNGGCCACCTGTGTTT 1373
QY 441 ValProSerLysValValGluValGlnGluIleProThrGlyGluProValCys 460
DB 1374 GTCCCACTCCCAAGTGTGTTGAGTCCAGAGGGCATCCCACTGGGGAGCGCTGTGTGT 1433
QY 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
DB 1434 GTCTACACTGCAGAAAGCCCTGCACAGAGAAATCAAAAGATCAGCTACCGCATCTCGAGA 1493
QY 481 AspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500

DB 1494 GACCCAGCAGGGTGGCTAGCCATGACCCAGACAGTGGCAGGTCCAGCTCTGGGCACC 1553
QY 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
DB 1554 CTCGACCGGTGAGATGAGCAGTTTGTGAGGAACAACATCTAAGAGTCAATGCTTGGCC 1613
QY 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuLeuAsp 540
DB 1614 ATGGCAATATGGAAAGCCCTCCACCACTGGCAGCGGAACCCCTTCTGTAACTGATTGAT 1673
QY 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
DB 1674 GTCAACGACCATGGCCCACTCCCTGAGCCCGTCAAGATCACCATCTGCAACCAAGCCCT 1733
QY 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
DB 1734 GTGCGCCAGCTGTGAACATCAAGCAAGGACCTGTCTCCACACCTCCCTTTCCAG 1793
QY 581 AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGluGlyAsp 600
DB 1794 GCGCAGCTCACAGATGATCTCAGACATCTACTGGACGCGCAGAGGTCAAAGAGAAAGTAC 1853
QY 601 ThrValValLeuSerLeuLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
DB 1854 ACAGTGGTCTTCTCCCTGAAGAAGTCTCTGAAGCAGGATACATATGACGTGCACCTTCT 1913
QY 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
DB 1914 CTGCTCTGACCATGGCAACAAGAGCAGCTGACGGTGATCAGGGCCACTGTGTGCACATGC 1973
QY 641 HisGlyHisValGluThrCysProGlyProTyrLysGlyGlyPheIleLeuProValLeu 660
DB 1974 CATGCCATGTGCAAACTCGCCCTGGAAAGAGAGGTTTCATCTCCTCCTGTGCTG 2033
QY 661 GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysLys 680
DB 2034 GGGGCTGTCTGGCTCTGCTGTCTCTGCTGTGCTGTGCTTTCTGTGTGAGAAGAG 2093
QY 681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
DB 2094 CGGAGATCAAGAGCGCCCTCTCTCTCCAGNAGATGACCCCGTGACACGCTCTTCTAC 2153
QY 701 TyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
DB 2154 TATGCGAAGAGGGGGTGGCGAAGAGCACAGGACTATGACATCACCCAGCTCCACCGA 2213
QY 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleLeuPro 740
DB 2214 GGTCTGGAGGCCAGCGCGGAGGTGTCTTCGCAATGACGTGGCACCACCATCATCCCG 2273
QY 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
DB 2274 ACACCCATGTACCGTCTTAGGCCAGCCAAACCCAGATGAAATCGGCACTTTTATATTGAG 2333
QY 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
DB 2334 AACCTGAAAGGGGCTTAAACACAGACCCACGCCCCCTTACGACACCTCTTGTGTTC 2393
QY 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
DB 2394 GACTATGAGGCGAGCGGCTCCGACCGCGGTCCCTGAGCTCCCTCACCTCTCCGCTCC 2453
QY 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGlySerArgPheLysLysLeuAla 820
DB 2454 GACCAAGACCAAGATTACGATTATCTGAACGAGTGGGCGAGCGGCTTCAAGAAGCTGGA 2513
QY 821 AspMetTyrGlyGlyGluAspAsp 829
DB 2514 GACATGTACGTGGCGGGAGGAGCGAC 2540

RESULT 4

US-10-058-270A-125

654 ATCTCCATCATCGTGAACCGACAGATGATCCACCAAGCCCAAGTTTATCCACGACACACCTTC 713
QY
221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
DB
714 CGAGGGAGTGTCTAGAGGGAGTCTTACCAGGTACTTCTGTGATGACGGTGAAGCCACA 773
QY
241 AspGluAspAlaIleThrThrTyAsnGlyValValAlaTySerIleHisSerGln 260
DB
774 CATGAGGATGATGCATCATCATCACTACATGAGGTGGTGTCTTACTCATCATAGCAA 833
QY
261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
DB
834 GAACCAAGAGGACCCACACAGCTCATGTTCACATTCACCGAGCAGACGACCATCAGC 893
QY
281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyThrLeuThrIleGlnAla 300
DB
894 GTCACTCCAGTGGCTGGACCGGGGAAAAGTCCCTGAGTACACATGACCATCCAGGCC 953
QY
301 ThrAspMetAspGlyAspGlySerThrThrThrAlaValAlaValGluIleLeuAsp 320
DB
954 ACAGACATGATGGGACCGGCTCCACCAACGACGTCAGTGGAGATCCTTGAT 1013
QY
321 AlaAspAsnAlaProMetPheAspProGlnIlystYrGluAlaHisValProGluAsn 340
DB
1014 GCCAATGCAATGCTCCCATGTTTACCCCGAGAGTACGAGGCCCATGTGCTGAGAA 1073
QY
341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
DB
1074 GCAGTGGGCCATGAGTGCAGAGCTGACGGTCACTGATCTGGACGCCCCCACTACCA 1133
QY
361 AlaTrpArgAlaThrTyLeuIleMetGlyAspAspGlyAspHisPheThrIleThr 380
DB
1134 CGTGGCGTGCACCTACCTATCATGGCGGGTACGACGCGGACCAATTTTACCATCACC 1193
QY
381 ThrHisProGluSerAsnGlnGlyLeuThrThrArgLysGlyLeuAspPheGluAla 400
DB
1194 ACCACCTTGAGACCAACCGGCATCTGACACACGAGGAGGGTTTGGATTTTGGGCC 1253
QY
401 LysAsnGlnHisThrLeuTyValGluValThrAsnGluAlaProPheValLeuLysLeu 420
DB
1254 AAAAACCAGACACACCTGTAGCTTGAAGTGAACCAACGAGGCCCTTTTGTGCTGAAGCTC 1313
QY
421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
DB
1314 CCACTTCCACGACCACTAGTGTCCAGTGGAGGATGTGATGAGGACACCTGTGTTT 1373
QY
441 ValProSerLysValValGluValGlnGluGlyLeuProThrGlyGluProValCys 460
DB
1374 GTCCACCCCTCCAAAGTCGTGAGGTCCAGGAGGCGCATCCCACTGGGGAGCCTGTGTGT 1433
QY
461 ValTyThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyArgIleLeuArg 480
DB
1434 GTCTACACTGCAGAGACCCCTGACAAAGAGAAATCAAAAGATCAGCTACCCGATCCTGAGA 1493
QY
481 AspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
DB
1494 GACCCAGCAGGCTGGCTAGCATGACCCACACAGTGGCGAGGTCAACAGTGTGGGCACC 1553
QY
501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyGluValMetValLeuAla 520
DB
1554 CTCGACCTGTAGGATGAGCATTTTGTGAGGAAACAAATCATATGAAGTCATGTTCTGGCC 1613
QY
521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAsp 540
DB
1614 ATGGACATGGAGCCCTCCACACCTGGCAGGGAACCCCTTCTGCTAATCACTGATTTGAT 1673
QY
541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
DB
1674 GTCAACGACCATGGCCCGCTGAGCCCGCTCAGATCACTCTGCAACCAAGCCCT 1733
QY
561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
DB
1734 GTGCGCCACGTGCTGAACATCAGCGAAGAGACCTGTCTCCCCACACCTCCCTTTCAG 1793

QY 581 AlaGlnLeuThrAspAspSerAspIleTyTrpThrAlaGluValAsnGluGluGlyAsp 600
DB 1794 GCCCAGCTCACAGATGACTCAGACATCTACTGGCGGAGAGGTCAACGAGGAAGGTGAC 1853
QY 601 ThrValValLeuSerLeuLysIlyPheLeuLysGlnAspThrTyAspValHisLeuSer 620
DB 1854 ACAGTGGTCTTGTCTCCCTGAAGAAGTCTCTGAAGCAGGATTCATATGACGTGCACCTTCT 1913
QY 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
DB 1914 CTGTCTGACCATGGACAAAGAGAGCTGACGGTATCAGGGCCACTGTGTGGACTGC 1973
QY 641 HisGlyHisValGluThrCysProGlyProTrpLysGlyGlyPheIleLeuProValLeu 660
DB 1974 CATGGCCATGTGAAACCTGCGCTGACCCCTGGAAGAGGAGGTTTTCATCTCCCTGTGCTG 2033
QY 661 GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuLeuValArgLysLys 680
DB 2034 GGGGCTGTCTGGCTCTGCTGTTCTCTCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2093
QY 681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTy 700
DB 2094 CGGAAGATCAAGGACCCCTCTACTCTCCAGAGATGACACCCGTCACACGCTCTTCTAC 2153
QY 701 TyrGlyGluGluGlyGlyGlyGluAspGlnAspTyArgPheIleThrGlnLeuHisArg 720
DB 2154 TATGCGAAGAGGGGGTGGCGAAGAGGACGAGACTATGACATCACCAGCTCCACCGA 2213
QY 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
DB 2214 GTCTGGAGGCCAGCGCGAGGGTGTCTCCGCAATGACGTGGCACCACCAACCATCCCG 2273
QY 741 ThrProMetTyArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
DB 2274 ACACCATGTATCCGTCTTAGCCAGCCCAACCCAGATGAATCGGCAACTTTTATTAATTGAG 2333
QY 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProTyArgPheThrLeuLeuValPhe 780
DB 2334 AACCTGAGGCGGCTTAACAGACCCCGCCGCTACGACACCCCTCTTGTGTTC 2393
QY 781 AspTyArgGlySerGlySerAspAlaSerLeuSerSerLeuThrSerSerAlaSer 800
DB 2394 GACTATGAGGCGAGCGGCTCCGACCGCGCTCCCTGAGCTCCCTCAGCTCCCTCCGCTCC 2453
QY 801 AspGlnAspGlnAspTyArgPheLeuAsnGluTrpGlySerArgPheLysLysLeuAla 820
DB 2454 GACCAAGACCAAGATTACGATTATCTGAACGAGTGGGCGAGCCGCTTCAAGAGCTGCA 2513
QY 821 AspMetTyArgGlyGlyGluAspAsp 829
DB 2514 GACATGTACGTGGCGGGGAGGACGAC 2540

RESULT 6

US-10-172-118-669
; Sequence 669, Application US/10172118
; Publication NO. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
; CURRENT APPLICATION NUMBER: US/10/172,118
; PRIOR APPLICATION NUMBER: 2002-06-14
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699

Db 1194 ACCACCCCTGAGAGCAACAGGGGATCTGACACCAAGAGGTTGGATTTTGGAGGCC 1253
Qy 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db 1254 AAAAACCAGCACACCTCTAGCTTGAAGTACCAACAGAGGCCCTTTTGTGTCGAAGCTC 1313
Qy 421 ProThrSerThrAlaThrIleValHisValGluAspValAsnGluAlaProValPhe 440
Db 1314 CCAACCTCCACAGCACCATAGTGGTCCACGTGGAGGATGTGAATGAGGACCTCTGTGTTT 1373
Qy 441 ValProProSerLysValValGluValGlnGluGlyLeuProThrGlyGluProValCys 460
Db 1374 GTCCCAACCTCCAAAGTGTGAGGTCCAGAGAGGACATCCCACTGGGGAGCCCTGTGTGT 1433
Qy 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db 1434 GTCTACACTGCAGAGACCTCTGACAGAGAGAAATCAAAAGATCAGCTACCGCATCTCTGAGA 1493
Qy 481 AspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db 1494 GACCCAGCAGAGGTGGCTAGCCATCGACCCAGACAGTGGGAGGTTCACAGCTGTGGGACC 1553
Qy 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db 1554 CTCGACCGTGAGGATGAGCAGTTTGTGAGGAACAACATCTATGAAGTCATGGTCTTGCC 1613
Qy 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAsp 540
Db 1614 ATGGACAATGAAGCCCTCCACCACTGGCAGCGGAAACCTCTCTGCTTAACACTGATTGAT 1673
Qy 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 1674 GTCAACGACCATGCGCCAGTCCCTGAGCCCGTCAGATCACCATCTGCAACCAAGCCCT 1733
Qy 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
Db 1734 GTGGCCACGTGTGTAACATCACGACAGACCTGTCTCCCAACACCTCCCTCTTCCAG 1793
Qy 581 AlaGlnLeuThrAspAspSerAspIleTyrThrAlaGluValAsnGluGlyAsp 600
Db 1794 GCCAGCTCAGATGACTAGACATCTACTGGACGCGAGGTCAACGAGGAAGGTGAC 1853
Qy 601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
Db 1854 ACAGTGTCTTGTCCCTGAAGAAGTCTCTGAAGCAGGATACATATGACGTGCACCTTCT 1913
Qy 621 LeuSerAspHisGlyAsnLysGluLeuThrValIleArgAlaThrValCysAspCys 640
Db 1914 CTGTCTGACCATGGCAACAAGAGCAGCTGACGTGATCAGGGCCACTGTGTCCGACTGC 1973
Qy 641 HisGlyHisValGluThrCysProGlyProTyrGlyGlyPheIleLeuProValLeu 660
Db 1974 CATGGCCATGTGCAAACTGCGCTGGACCTGGAAAGAGGTTTCATCTCCCTCTGTGCTG 2033
Qy 661 GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysLys 680
Db 2034 GGGGCTCTCTGGCTCTGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2093
Qy 681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
Db 2094 CGGAGATCAAGAGCCCTCTCTACTCCAGAGATGACACCGTGACACAGTCTCTCTAC 2153
Qy 701 TyrGlyGluGlyGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
Db 2154 TATGGCAAGAGGGGGGTGGCAAGAGGACAGGACTATGACATCACCAGCTCCACCGA 2213
Qy 721 GlyLeuAlaArgProGluValValValLeuArgAsnAspValAlaProThrIleIlePro 740
Db 2214 GGTCTGAGGCCAGCGCGAGGTGTCTCCGCAATGACGTGGCACAACCACTATCCCG 2273
Qy 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760

Db 2274 ACACCATGTACCGTCTAGCGGACCAACCCAGATGAATCGCAACTTTTATTAATTGAG 2333
Qy 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
Db 2334 AACCTGAAGGGGCTAACACAGACCCCAAGCCCGCCCTACGACACCTCTTGTGTTTC 2393
Qy 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
Db 2394 GACTATGAGGCGAGCGGCTCCGACGCGGCTCCCTGAGCTCCCTCACCTCTCCGCTCC 2453
Qy 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTyrGlySerArgPheLysLysLeuAla 820
Db 2454 GACCAAGACCAAGATTACGTTATCTGAACGATGGGGCAGCGCTTCGAAGAAGCTGGCA 2513
Qy 821 AspMetTyrGlyGlyGlyGluAspAsp 829
Db 2514 GACATGTACCGTGGCGGGAGGACGAC 2540

RESULT 8
US-10-158-123-2
; Sequence 2, Application US/10158123
; Publication No. US20030194406A1
; GENERAL INFORMATION:
; APPLICANT: Reinhard, Christoph
; APPLICANT: Klinger, Julie
; APPLICANT: Jefferson, Ann
; APPLICANT: Escobedo, Jaime
; APPLICANT: Randazzo, Filippo
; APPLICANT: Winter, Jill
; APPLICANT: Goodson, Robert
; APPLICANT: Qi, Weimin
; TITLE OF INVENTION: P-Cadherin as a Target for Anti-Cancer
; TITLE OF INVENTION: Therapy
; FILE REFERENCE: 35784/258994
; CURRENT APPLICATION NUMBER: US/10/158,123
; CURRENT FILING DATE: 2003-02-07
; PRIOR APPLICATION NUMBER: 60/294,225
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 3171
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-158-123-2

Alignment Scores:
Pred. No.: 0 Length: 3171
Score: 4369.00 Matches: 829
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 15 Gaps: 0

US-09-916-849A-1 (1-829) x US-10-158-123-2 (1-3171)

Qy 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuValCysTrpLeuGln 20
Db 54 ATGGGGCTCCCTCGTGGACCTTCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 113
Qy 21 CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
Db 114 TCGCGGGCTCCGAGCCCTCGCGGCTCTCTCAAGGAGGCTGAAGTACCTTGGAGCGC 173
Qy 41 GlyGlyAlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
Db 174 GAGAGCGGGAGCAGGACCCCGGCGAGCGCTGGGGAAGATTTCNTGGGCTGCCCTGGG 233
Qy 61 GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyIleThr 80
Db 234 CAAGAGCCAGCTCTCTTTAGCATTGATATGATGATGATGATGATGATGATGATGATGATG 293
Qy 81 ValGlnGluArgArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100

294 GTCCAGGAAAGAGTCACTGAAGAAAGAAATCCATTGAAGATCTTCCCATCCAAACGT 353
 101 IleLeuArgHisLysAtcAspTTPValValAlaProIleSerValProGluAsnGly 120
 354 ATCTTACAGACACAAAGAGATTGGTGGTGGCTCAATATCTGCTCCCTGAATGGC 413
 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
 414 AAGGGTCCCTTCCCCAGAGACTGAATCAGCTCAAGTCAATTAAGATAGAGACACCAAG 473
 141 IlePheTyrSerIleThrGlyProGlyValAlaAspSerProGluGlyValPheAlaVal 160
 474 ATTTTCTACAGCATCAGGGGCCCGGGGCGACAGCCCCCTGAGGGTGTCTTCGGTGTGA 533
 161 GlnLysGluThrGlyTTPLeuLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys 180
 534 GAGAGGAGACAGGCTGGTGTGTGTAATAAGCCACTGACCGGGAGGAGATTGCCAAG 593
 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
 594 TATGAGCTCTTTGGCCACGCTGTGTACAGAAATGGTGCCTCAGTGGAGGACCCCATGAAC 653
 201 IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
 654 ATCTCCATCATCGTACCGACACAGATGATGACCAAGCCCAAGTTTACCCAGGACACCTTC 713
 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
 714 CGAGGAGTGTCTTAGAGGAGTCTTACCAGGTACTTCTGTGTGTCAGGTGCACAGCCACA 773
 241 AspGluAspAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln 260
 774 GATGAGGATGATGCCATCTACCTACCTACATGCGGTGGTGTCTTACTTCATCCATAGCCAA 833
 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
 834 GAACCAAGGACCCACACGACTCATGTTCACATTCACCGAGCAGACGACCATCAGC 893
 281 VallIleSerSerGlyLeuAspArgGlnLysValProGluTyrThrLeuThrIleGlnAla 300
 894 GTCATCTCCAGTGGCTCGACCGGAAAGTCTCCTGAGTACACACTGACATCCAGGCC 953
 301 ThrAspMetAspGlyAspGlySerThrThrAlaValAlaValAlaValGluIleLeuAsp 320
 954 ACAGACATGGATGGGACCGCTCCACCAACCGCAGTGGCGAGTAGTGGAGATCCTTGAT 1013
 321 AlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
 1014 GCCAATGACATGTCTCCCATGTTTGAACCCACAGAGTACGAGGCCCATGTGCTCGAGAAT 1073
 341 AlaValGlyHisGluValGluArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
 1074 GCAGTGGGCCCATGAGTGCAGAGGCTGACGCTACTGATCTGGAGCCCCCACTCACCA 1133
 361 AlaThrArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
 1134 GGTGGCGTGCACCTATCATATGCGCGGTGACGAGGCCCATTTTACCATCACC 1193
 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
 1194 ACCCACTTGAGAGCAACAGGSCATCTGTACACACAGAGAGGTTTGGATTTTGGGCC 1253
 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
 1254 AAAAACCAGCACACCTGTACGTGAAGTGAACCAACAGAGGCCCTTTTGTGCTGAAGCTC 1313
 421 ProThrSerThrAlaThrIleValHisValGluAspValAsnGluAlaProValPhe 440
 1314 CCAACCTCCAGGACCATAGTGTGTCAGCTGGAGGATGGAATGAGGACCTGTGTTT 1373
 441 ValProProSerLysValValGluValGlnGluIleProThrGlyGluProValCys 460

1374 GTCCACCTCCAAAGTCGTTGAGGTCCAGGAGGGCATCCCACTGGGGAGCCTGTGTGT 1433
 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
 1434 GTCTACACTGCAGAGACACCTGCAGAGGAGAAATCAAAAGATCAGTACCCGATCCTGAGA 1493
 481 AspProAlaGlyTTPLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
 1494 GACCAGCAGGGTGGCTAGCCATGGACCCAGACAGTGGGACAGGTACAGCTGTGGGCACC 1553
 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
 1554 CTCGACCGTGGATGAGCAGTTTGTGAGGAACAACATCTATGAAGTCATGCTTGGCC 1613
 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuLeuIleAsp 540
 1614 ATGGCAATGGAAGCCCTCCACCATCTGGCAGGGAACCCCTTCTGCTAACACTGATTGAT 1673
 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
 1674 GTCAAGCACCATTGGCCAGTCCCTGAGCCCCGTGAGATCACCATCTGCAACCAAGCCCT 1733
 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
 1734 GTGCGCCACGTGCTGAACATCAGGACAAAGACCTGTCTCCCAACACCTCCCTTTCAG 1793
 581 AlaGlnLeuThrAspAspSerAspIleTyrTTPThrAlaGluValAsnGluGluGlyAsp 600
 1794 GCCCAGCTCAGATGACTCAGACATCTACTGGAGCGGACAGAGTCAACGAGGAAGTGAC 1853
 601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
 1854 ACAGTGGTCTTCTCCCTGAAGAAGTCTCTGAAGCAGGATACATATGACGTGCACCTTCT 1913
 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
 1914 CTGCTGACCATGGCAACAGAGAGCTGACGTGATCAGGGCCACTGTGTGGACTGC 1973
 641 HisGlyHisValGluThrCysProGlyProTTPLysGlyGlyPheIleLeuProValLeu 660
 1974 CATGGCATGTGCAAACTGCCCTCGAACCTGGAAAGGAGGTTTCATCTCCCTGTGTGTG 2033
 661 GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysLys 680
 2034 GGGGCTGTCTGGCTGTGTCT 2093
 681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
 2094 CGGAAGATCAAGAGGCCCTCTCTACTCCAGAAAGATGACACCCGTGACACAGCTCTCTAC 2153
 701 TyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnHisArg 720
 2154 TATGGCAAGAGGGGGTGGCAAGAGGACCAAGGACTATGACATCACCCAGCTCCACCGA 2213
 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
 2214 GGTCTGAGGCCAGGCCGCGAGGTGTCTCTCCGATGACGTGGCACCAACCATCTCCCG 2273
 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
 2274 ACACCCATGTACCGTCTAGGCGCAGCAACCCAGATGAATCGCAACTTATATAATTGAG 2333
 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
 2334 AACCTGAAGCGGCTTAACACAGCCCCACAGCCCGCCCTACGACACCTCTGTGTGTT 2393
 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
 2394 GACTATGAGGCGAGCGCTCCGACCGCGTCCCTGAGCTCCCTCAGCTCCCTCCCGCTCC 2453
 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTTPGlySerArgPheLysLysLeuAla 820
 2454 GACCAAGACCAAGATTACGATTATCTGACAGAGTGGGCGAGCGCTTCAAGAGAGTGGA 2513

Db	174	GGAGCGCGGAGCAGAGCCGGCCAGCGCTGGGGAAAGTAATTATGGCTGCCCTGGG	233
Qy	61	GInGluProAlaLeuPheSerThrAspIenAspPheThrValArgAsnGlyGluThr	80
Db	234	CAAGAGCCAGCTCTGTTTAGCAGCTGATATGATGATCTTCACTGTGGGAATGGCAGACA	293
Qy	81	ValGInGluArgArgSerLeuLeuGlyGluArgAsnProLeuLysIlePheProSerLysArg	100
Db	294	GTCCAGGAAAGAGGTCACTGAAGAAAGGAATCCATTGAAGATCTTCCATCCAAACGT	353
Qy	101	IleLeuArgArgHisLysArgAspTrpValValAlaProIleSerValProGluAsnGly	120
Db	354	ATCTTACCAAGACACAAGAGAGATTGGTGGTGTCTCAATATCTGTCCCTGAAATGCG	413
Qy	121	LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys	140
Db	414	AAGGGTCCCTTCCCCCAGAGACTGAATCAGTCTCAAGTCTAATAAGATAGAGACACCAAG	473
Qy	141	IlePheTySerIleThrGlyProGlyAlaAspSerProProGluGlyValPheAlaVal	160
Db	474	ATTCTTACAGCATCACGGGCGCGGGCAGACAGCCCGCTGAGGGTGTCTTCCTCTGTA	533
Qy	161	GluLysGluThrGlyTrpLeuLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys	180
Db	534	GAGAAGGAGACAGCGTGGTGTGTGAATAAGCCATCGGACCGGGAGAGATTGCCAAG	593
Qy	181	TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn	200
Db	594	TATGAGTCTTTTGCCACAGCTGTGTACAGAAATGGTGCTCAGTGGAGACCCCATGAAC	653
Qy	201	IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe	220
Db	654	ATCTCCATCATCGTACCGACCAAGATGACCACAAAGCCCAAGTTTACCAGGACACCTTC	713
Qy	221	ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr	240
Db	714	CGAGGGAGTGTCTTAGAGGGAGTCCCTACAGGTACTTCTGTGATGCAGGTGACAGCCACA	773
Qy	241	AspGluAspAspAlaIleTyThrTyAsnGlyValValAlaTySerIleHisSerGln	260
Db	774	GATGAGATGATGCCATCTACACCTACATGGGGTGGTGTCTTACTTCATCCATGACCAA	833
Qy	261	GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer	280
Db	834	GAACCAAGACCCACACAGCACTCATGTTCCAAATTCACGGAGCAGGACCATCAGC	893
Qy	281	ValIleSerSerGlyLeuAspArgGluLysValProGluTyThrLeuThrIleGlnAla	300
Db	894	GTCATCTCCAGTGGCTGGACCGGAAAAAGTCCCTGAGTACACACTGACCATCAGGCC	953
Qy	301	ThrAspMetAspGlyAspGlySerThrThrThraValaValaValGluIleLeuAsp	320
Db	954	ACAGACATGATGGGAGCGCTCCACCACACCGCAGTGGCAGTAGTGGAGATCCTTGAT	1013
Qy	321	AlaAsnAspAsnAlaProMetPheAspProGlnLysTyThrGluAlaHisValProGluAsn	340
Db	1014	GCCAAATGACAATGCTCCCATGTTTGACCCCAAGAGTAGAGGGCCCATGTGCTTGAGAT	1073
Qy	341	AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro	360
Db	1074	GCASTGGGCCATGAGGTGACAGCGCTGACCGTCACTGATCTGAGCGCCCACTCACCA	1133
Qy	361	AlaTrpArgAlaThrTyLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr	380
Db	1134	GCGTGGCGTGCCACCTACCTTATCATGGGGGTGACGACGGGGACCAATTTTACATCACC	1193
Qy	381	ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla	400
Db	1194	ACCCACCTGTAGAGCAACACGGGCATCCTGTACAAACAGGAAGGGTGTGGATTTTGAGGCC	1253
Qy	401	LysAsnGlnHisThrLeuTyValGluValThrAsnGluAlaProPheValLeuLysLeu	420

Db 1254 AAAAAACGACACACCCCTGACCTGAAGTGAACCAAGAGCCCTTTTGTCTGACGCTC 1313
Qy ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
Db 1314 CCAACCTCCACACACCCACCATAGTGTCCACGTGGAGGATGTAATGAGCACCTGTGT 1373
Qy ValProProSerIysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
Db 1374 GTCCCACTCCCAACCTGCTGAGTCCAGGAGGACATCCCACTGGGAGCCCTGTGTGT 1433
Qy ValTyThrAlaGluAspProAspIysGluAsnGlnLysIleSerTyArgIleLeuArg 480
Db 1434 GTCTACATGCAGAACCCCTGACAGGAGATCAAAAGATCAGCTACCGCATCTCGAGA 1493
Qy AspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db 1494 GACCCAGCAGGGTGGTGTAGCCATGACACCCAGACAGTGGGCAAGTCCACAGCTGTGGGCACC 1553
Qy LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db 1554 CTCGACGTCGAGGATGAGCAGTTTGTGAGGAAACAACATCTATGAATCATGTCTTGCC 1613
Qy MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAsp 540
Db 1614 ATGGACAAATGGAGCCCTCCCACTGTCGACGGGAACCTTCTGCTAACCACTGATTGAT 1673
Qy ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 1674 GTCAACGACCATGGCCCTGCTGAGCCCTGTCAGATCACCATCTGCAACCAAGCCCT 1733
Qy ValArgHisValLeuAsnIleThrAspIysAspLeuSerProHisThrSerProPheGln 580
Db 1734 GTGGCGACATGCTGAACATCAAGCAAGGACCTGTCTCCCAACACCTCCCTTTCCAG 1793
Qy AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGlyAsp 600
Db 1794 GCCCAGCTCACAGATGACTCAGACATCTACTGACGCGCAGAGGTCAACGAGGAAGTGAC 1853
Qy ThrValValLeuSerLeuIysIysPheLeuIysGlnAspThrTyrAspValHisLeuSer 620
Db 1854 ACAGTGTCTTGTCCCTGAAGAGTTCCTGACGAGGATACATATGATGACGTGCACCTTCT 1913
Qy LeuSerAspHisGlyAsnIysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
Db 1914 CTGTCTCAGCATGGCAACAGACAGCAGTGCAGGTGATCGGCCCATCTGTGTGACATGC 1973
Qy HisGlyHisValGluThrCysProGlyProTrpIysGlyIlePheLeuLeuProValLeu 660
Db 1974 CATGGCCATGTGCAAACTGGCCCTGGACCCCTGGAAGAGGATTTTCATCCTCCCTGCTG 2033
Qy GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysLys 680
Db 2034 GGGGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2093
Qy ArgLysIleIysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
Db 2094 CGGAAGATCAAGAGCCCTCTCTCTCCCAAGATGACACCCGTCGACACGCTCTCTCTAC 2153
Qy TyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
Db 2154 TATGGCGAAGAGGGGGTGGCGAAGAGGACGACGATGATGATCACCAGCTCCACCGA 2213
Qy GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleLeuPro 740
Db 2214 GGTCTGGAGCCAGCCCGAGGTGCTTCTCCGCAATGACGTGGCACCACCAACATCATCCG 2273
Qy ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleLeuGlu 760
Db 2274 ACACCCATCTACCGTCTTAGCCACCCACCCAGATGAATCGGCACTTATATATGAG 2333
Qy AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
Db 2334 AACCTGAAGGGGGCTTAACACAGACCCCAAGCCCGCCCTACGACACCCCTCTTGTGTTC 2393

Qy 781 AspTyrGluGlySerGlySerAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
Db 2394 GACTATGAGGCGAGCGCTCCGACGCGCTCCCTGAGTCCCTCACCTCTCCGCTCC 2453
Qy 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGlySerArgPheIysLeuAla 820
Db 2454 GACCAAGACCAAGATTACATTATCTGAACGAGTGGGCGACCGCTTCAAGAAGCTGGCA 2513
Qy 821 AspMetTyrGlyGlyGlyGluAspAsp 829
Db 2514 GACATGTACGCTGGCGGGGAGGACGAC 2540
RESULT 10
US-10-295-027-799
; Sequence 799, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 799
; LENGTH: 3171
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-295-027-799
Alignment Scores:
Pred. No.: 0 Length: 3171
Score: 4369.00 Matches: 829
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 16 Gaps: 0
US-09-916-849A-1 (1-829) x US-10-295-027-799 (1-3171)
Qy 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuGlnValCysTrpLeuGln 20

[illegible]

1134	Db	CGCTGGCGGTGCACCATCTACCTTATCATGGCGGGTACGACGAGGGGACCATTTTTACCATCACCC	1193
381	Qy	ThrHisProGluSerAsnGlnGlyIleIleuThrThrArgIysGlyLeuAspPheGluAla	400
1194	Db	ACCCACCCCTGAGAGCAACCAAGGGCATCCTGACAAACAGGAGGGTTTGGATTTTTCAGGCC	1253
401	Qy	LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu	420
1254	Db	AAAAACAGCACACCCCTGTAGTTTGAAGTGACCAACGAGGCCCTTTTGTGCTGAAGCTC	1313
421	Qy	ProThrSerThrAlaThrIleValValHisValGluAspValIleAsnGluAlaProValPhe	440
1314	Db	CCAACTCCACAGCCACCATAGTGGTCCACGTGGAGGATGTGAATAGGCACCTGTGT	1373
441	Qy	ValProProSerLysValValGluValGlnGlnGlyIleProThrGlyGluProValCys	460
1374	Db	GTCCACCCCTCCAAAGTCGTTGAGTCCAGAGGGGCATCCCCACTGGGAGAGCTGTGTGT	1433
461	Qy	ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg	480
1434	Db	GTCTACACTGCAGAGACCCCTGCAAGGAGAAATCAAAGAATCAGCTACCGCATCTTGAGA	1493
481	Qy	AspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr	500
1494	Db	GACCCAGCAGGTGGCTAGCCATGGACCCACAGTGGCGAGTCAAGCTGTGGGCACC	1553
501	Qy	LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla	520
1554	Db	CTCGACCGTGCAGGATGAGCAGTTTGTGAGGAACAACATCTATGAAGTCATGCTTGGCC	1613
521	Qy	MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrIleLeuAsp	540
1614	Db	ATGGACATATGGAGGCCCTCCACATCGGCACGGGAACCCCTCTGCTAACACTGATTGAT	1673
541	Qy	ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro	560
1674	Db	GTCAACGACCATGGCCCGAGTCCCTGAGCCCGTCAGATCACCATCTTGCAACCAAGCCCT	1733
561	Qy	ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln	580
1734	Db	GTGGCCACATGCTGTAACATCAGCGACAGGACCTGTCTCCCCACACCTCCCTTCCAG	1793
581	Qy	AlaGlnLeuThrAspAspSerAspIleTyrTyrThrAlaGluValAsnGluGluValAsp	600
1794	Db	GCCAGCTCACATGACTCAGACATCTACTGGACGGCAGAGGTCAACGAGGAAGTGAC	1853
601	Qy	ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer	620
1854	Db	ACAGTGGTCTGTGCCCTGAGAGAGTTCCCTGAAGCAGGATACATATGACGTGCACCTTCT	1913
621	Qy	LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys	640
1914	Db	CTGTCTGACCATGGCAACAAGAGCAGCTGACGGTGATCAGGGCCACTGTGTGCGAGTCG	1973
641	Qy	HisGlyHisValGluThrCysProGlyProTyrLysGlyGlyPheIleLeuProValLeu	660
1974	Db	CATGGCCATGTGAAACCTGCCCTGGACCCCTGAAAGGAGGTTTTCATCTCCCTGTGCTG	2033
661	Qy	GlyAlaValIleAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysLys	680
2034	Db	GGGGCTGTCTGGCTCTGTGTCTCTCTGCTGGTGCTCTCTTTTGTGTGTGAAGAAAG	2093
681	Qy	ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr	700
2094	Db	CGAAGATCAAGAGAGCCCTCTACTTCCCAAGAAGATGACACCGGTGACACGCTCTCTAC	2153
701	Qy	TyrGlyGluGlnGlyGlyGlyGluAspGlnAspTyrAspIleThrClnIleuHisArg	720
2154	Db	TATGGCGAAGAGGGGGTGCGAAGAGGACAGGACTATGACATCACCAGCTCCACCGA	2213
721	Qy	GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro	740
2214	Db	GGTCTGGAGCCAGGCCGAGGTGGTTCTTCCCAATGACGTGGCACCAACCATATCCCG	2273

Db 1014 GCCAATGCAATGCTCCCATGTTTGACCCCCAGAGTACGAGGCCCATGTGCTGAGAAAT 1073
Qy 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
Db 1074 GCAGTGGCCCATGAGGTGACAGGCTGACGCTGCTACTGATCTGGAGCGCCCACTACCA 1133
Qy 361 AlaTrpArgAlaThrTyrLeuIleMetGlyGlyAspGlyAspHisPheThrIleThr 380
Db 1134 GCGTGGCGTCCACCTACCTATCATGGCGGTGACGACGGGACCACTTTTACCATCACC 1193
Qy 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
Db 1194 ACCACCCCTGAGACCAACAGGCGATCCTGCAACACAGGAGGCTTTGGATTTTGAGGCC 1253
Qy 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db 1254 AAAACACGACACACCTGTAGTGAAGTACCAAGGCGCCCTTTTGTGCTGAGACTC 1313
Qy 421 ProThrSerThrAlaThrIleValHisValGluAspValAsnGluAlaProValPhe 440
Db 1314 CCAACCTCCACAGCCACCATAGTGGTCCACGTGAGGATGTGAATGAGGCACCTGTGTT 1373
Qy 441 ValProProLysValValGluValGlnGlyIleProThrGlyGluProValCys 460
Db 1374 GTCCACCCCTCCAAAGTGTGAGGTCCAGAGGCGCATCCCACTGGGGAGCCTGTGTGT 1433
Qy 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db 1434 GTCTACACTGCAGAGAACCTCTGACAGAGAAATCAAAAGATCAGTACCGCATCCTGAGA 1493
Qy 481 AspProIaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db 1494 GACCCACAGCGTGGCTAGCCATGACATGACCCAGACAGAGTGGGAGGTACAGCTGTGGGCACC 1553
Qy 501 LeuAspArgGluAspGluGlnPheValArgAsnIleTyrGluValMetValLeuAla 520
Db 1554 CTCGACCGTGGAGTACAGCTTTGTGAGGACAAACATCTATGATGATCACTGTGCTGGCC 1613
Qy 521 MetAspAsnGlySerProProThrThrThrGlyThrLeuLeuLeuThrLeuIleAsp 540
Db 1614 ATGGACAATGAGACCCCTCCACCACCTGCGGACGGGAACCCCTCTGCTTAACACTGATGAT 1673
Qy 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 1674 GTCAACGACCATGCCCGAGTCCCTGAGCCCGTCAGATCACCATCTGCAACCAAGCCCT 1733
Qy 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
Db 1734 GTGGCCACGCTGCTGAACATCAGGACACAGGACCTGTCTCCACACCTCCCTTTCCAG 1793
Qy 581 AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGlyAsp 600
Db 1794 GCCCAGCTCAGATGATGATCAGACATCTACTGGACGCGAGGTCAACGAGGAAGTGAC 1853
Qy 601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
Db 1854 ACAGTGTCTGTGCTTGGAGAGTCTCTGACGAGATACATATGACGTGCACCTTTCT 1913
Qy 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
Db 1914 CTGTCTGACCATGCAACAAAGAGCAGCTGACGGTGTATCAGGGCCACTGTGTGCGACTGC 1973
Qy 641 HisGlyHisValGluThrCysProGlyProTyrPheGlyPheIleLeuProValLeu 660
Db 1974 CATGGCCATGTGAAACCTCCCTGGACCTCGAAAGGAGGTTCATCTCTCCCTGTGCTG 2033
Qy 661 GlyAlaValLeuAlaLeuLeuPheLeuLeuLeuValLeuLeuLeuValArgLysLys 680
Db 2034 GGGGCTGTCTGCTGTCTGCTTCTCTCTCTGCTGCTGCTTTTGTGTGTGAGAAAGAG 2093
Qy 681 ArgLysIleLysGluProLeuLeuLeuProGluAspThrArgAspAsnValPheTyr 700
Db 2094 CGGAAGATCAAGGAGGCCCTCTCTACTCCCAAGAGATGACACCCGTGACAAACGCTTCTAC 2153

Qy 701 TyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
Db 2154 TATGGCGAAGAGGGGGGTGCGAAGAGGACCAAGGACTATGACATCACCCAGCTCCACCGA 2213
Qy 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
Db 2214 GGTCTGGAGGCCAGCGCGGAGGTGTCTCCGCAATGAGTGGCACCAACCATCATCCG 2273
Qy 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
Db 2274 ACACCCATGATCGCTAGGCGAGCCACCCAGATGAAATCGCAACTTTTATAATTGAG 2333
Qy 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
Db 2334 AACCTGAAGCGGCTTAACACAGCCCGCCGCTACACACCTCTTGTGTTC 2393
Qy 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
Db 2394 GACTATGAGGCGAGCGGCTCCGACGCGGTCCCTGAGCTCCCTCACCTCTCCGCTCC 2453
Qy 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTyrGlySerArgPheLysLeuAla 820
Db 2454 GACCAAGACCAAGATTACGATTATCTGAACGAGTGGGGGCGCTTCAAGAGCTGGCA 2513
Qy 821 AspMetTyrGlyGlyGlyGluAspAsp 829
Db 2514 GACATGTACGTTGGCGGGGAGGACGAC 2540

RESULT 12
US-10-295-027-895
; Sequence 895, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of diagnosis of Cancer, Compositions and
; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 895

894	Db	GTCAATCTCCAGTGGCCCTGGACCGGGAAAAAGTCCCTCTAGTACACACTGACCATCCAGGCC	953
301	Qy	ThrAspMetAspGlyAspGlySerThrThrThrAlaValaValValGluIleLeuAsp	320
954	Db	ACAGACATGGATGGGACGGCTCCACCACCGCAGTGGCAGTAGTGGAGATCCTTGAT	1013
321	Qy	AlaAspAspAsnAlaProMetPheAspProGlnIlySerGluAlaHisValProGluAsn	340
1014	Db	GCCAAATGACAATGCTCCCATGTTTGACCCCCAGAGTACAGGCCCCATGTGCTGAGAAT	1073
341	Qy	AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro	360
1074	Db	GCAGTGGGCCATGAGGTGCGAGAGGCTGACGGTCACTGATCTGTGACGCCCCCACTCACCA	1133
361	Qy	AlaTrpArgAlaThrTyrrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr	380
1134	Db	GGCTGGCGTGCCACCTTACCTTATCATGGGCGGTGACGACGGGACCATTTTACATCACC	1193
381	Qy	ThrHisProGlnSerAsnGlnGlyIleLeuThrThrArgIysGlyLeuAspPheGluAla	400
1194	Db	ACCCACCTTGAGAGCAACACGAGGCATCTCTGACAAACAGGAAGGTTTGATTTTGAGGCC	1253
401	Qy	LysAsnGlnHisThrIleuTyrrValGluValThrAsnGluAlaProPheValLeuLysLeu	420
1254	Db	AAAAACACGACACACCTGTGATCGTTGAAGTGACCAACGAGGCCCTTTTGTGCTGAAGCTC	1313
421	Qy	ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe	440
1314	Db	CCAACTCCACACGCCACCATAGTGGTCCAGTGGAGGATCTGATGAGGCACCTGTGTTT	1373
441	Qy	ValProProSerLysValValGluValGlnGlyIleProThrGlyGluProValCys	460
1374	Db	GTCCACCCCTCCAAAGTCGTGTAGGTCCAGAGGGCATCCCCACTGGGGAGCCTGTGTGT	1433
461	Qy	ValTyrrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrrArgIleLeuArg	480
1434	Db	GTCTACACTGCAGAAGACCTCTGACAGGAGAAATCAAAAGATCAGTACCGCATCTTGAGA	1493
481	Qy	AspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr	500
1494	Db	GACCCAGCAGGGTGGCTAGCCATGGACCCACAGACAGTGGGAGGTTCACAGCTGGGGCACC	1553
501	Qy	LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrrGluValMetValLeuAla	520
1554	Db	CTGCACCGTGAAGATGAGCAGTTGTGAGGAAACAACATCTATGAAGTCATGGCTCTGGCC	1613
521	Qy	MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAsp	540
1614	Db	ATGGACAAATGGAAGCCCTCCACACCTGGCACGGGAACCCCTTCGTCTAAACACTGATTGAT	1673
541	Qy	ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro	560
1674	Db	GTCAACACCATGGCCAGTCCCTGAGCCCGCTGAGATCCACATCTCTGCAACCAAGCCCT	1733
561	Qy	ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln	580
1734	Db	GTGGCCCACTGCTGTAACATCACGGAACAAGACCTGTCTCCACACCTCCCCCTTCCAG	1793
581	Qy	AlaGlnLeuThrAspAspSerAspIleTyrrTrpThrAlaGluValAsnGluGlyAsp	600
1794	Db	GCCCAGCTCAGATGATCTCAGACATCTACTGGACGGCAGAGGTCAACAGGAAGGTGAC	1853
601	Qy	ThrValValLeuSerLeuLysPheLeuLysGlnAspThrTyrrAspValHisLeuSer	620
1854	Db	ACAGTGTCTTGTCCTGAAGAGATTCTCTGAAGCAGGATCATATGACGTGCACCTTCT	1913
621	Qy	LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys	640
1914	Db	CTGTCTGACCATGGCAACAAGAGACAGCTGACGGTGTATCAGGGCCATGTGTGCGACTGC	1973
641	Qy	HisGlyHisValGluThrCysProGlyProTrpLysGlyGlyPheIleLeuProValLeu	660
1974	Db	CATGGCCATGTGCAAACTGCCCTGGACCCCTGGAAAGGAGGTTTTCATCTCCCTGTGCTG	2033

QY 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgIysGlyLeuAspPheGluAla 400
DB 1194 ACCCACCTGTAGAGCAACAGGAGTCTGTGCAACACAGGAGGTTTGGATTGTAGGCC 1253
QY 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
DB 1254 AAAAACCCAGCAACCCCTGTGACCTTGTGAGTGCACCAACGAGGCCCTTTGTGCTGAGCTC 1313
QY 421 ProThrSerThrAlaThrIleValHisValGluAspValAsnGluAlaProValPhe 440
DB 1314 CCAACCTCCACAGGCACCATAGTGTCCACGTGGAGGATGGAATGAGGCACCTGTGTTT 1373
QY 441 ValProProSerLysValValGluValGlnGlyIleProThrGlyGluProValCys 460
DB 1374 GTCCACACCTCCAAAGTCGTTGAGGTCCAGAGGAGCATCCCACTGGGAGCCTGTGTGT 1433
QY 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
DB 1434 GTCTACACTGCAGAAGACCTTGACAGGAGGAATCAAAAGATCAGCTACCGCATCCTGAGA 1493
QY 481 AspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
DB 1494 GACCCAGCAGGCTGTGCTAGCATGACCCAGACAGTGGCAGGTCAAGCTGTGGGCACC 1553
QY 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
DB 1554 CTCGACCTGTAGGATGAGCAGTTGTGAGGACACATCTATGAGTCATGGTCTTGCC 1613
QY 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAsp 540
DB 1614 ATGGACAATGGAAGCCCTCCACCACTGGCAGCGGAAACCTTCTGTCTAACCATGATGAT 1673
QY 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
DB 1674 GTCAACGACCATGGCCCGAGTCCCTGAGCCCGCTGAGATCAACATCTGCAACCAAGCCCT 1733
QY 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
DB 1734 GTGGCCACGCTGTCAACATCAGGACCAAGACCTGTCTCCACACACTCCCTTTCAG 1793
QY 581 AlaGlnLeuThrAspAspSerAspIleTyrThrAlaGluValAsnGluGlyAsp 600
DB 1794 GCCCAGCTCACAGATGACTCAGACATCTACTGAGCGGAGGTCACAGGAGAGGTGAC 1853
QY 601 ThrValValLeuSerLeuLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
DB 1854 ACAGTGGTCTTGCTCCCTGAAGAGTCTCTGAAGCAGGATACATATGACGTGACCTTTCT 1913
QY 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
DB 1914 CTGTCTGACCATGGCAACAAAGACAGCTGACGCTGATCAGGCGCATGTGTGCGACTGC 1973
QY 641 HisGlyHisValGluThrCysProGlyProTyrLysGlyGlyPheIleLeuProValLeu 660
DB 1974 CATGGCCATGTGCAAACTCGCTGGACCTTGGAAGAGGTTTCACTCCCTGCTGCTG 2033
QY 661 GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysLys 680
DB 2034 GGGGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2093
QY 681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
DB 2094 CGAAGATCAGAGGACCCCTCTACTCCAGAGATGACACCCGTGACACCTGCTCTTCTAC 2153
QY 701 TyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
DB 2154 TATGCCGAAGAGGGGGTGGCGAAGAGGACAGGACATGACATCAACCCAGCTCCACCGA 2213
QY 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleLeuPro 740
DB 2214 GGTCTGGAGGCGAGCGGAGTGTCTCCGCAATGACGTGGCACCACCATCATCTCCG 2273
QY 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleGlu 760

DB 2274 ACACCATGTACCTGCTTAGGCCACCCACAGATGAATCGCAACTTATTAATGAG 2333
QY 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProTyrAspThrLeuLeuValPhe 780
DB 2334 AACCTGAAGCGGCTAACACACAGCCACAGCCCGCCTACGACACCTCTTGGTGTTC 2393
QY 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
DB 2394 GACTATGAGGCGAGCGGCTCCAGCCCGCTCCCTGAGCTCCCTCAGCTCTCCGCTCC 2453
QY 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGlnIleTyrGlySerArgPheLysLeuAla 820
DB 2454 GACCAAGACCAAGATTACGATTATCTGAACGAGTGGGCGAGCGCTTCAAGAAGTGGCA 2513
QY 821 AspMetTyrGlyGlyGlyGluAspAsp 829
DB 2514 GACATGTACGTGGCGGGGAGGACGAC 2540
RESULT 14
US-10-775-920-188
; Sequence 188, Application US/10775920
; Publication No. US2004017544A1
; GENERAL INFORMATION:
; APPLICANT: Mergen Ltd
; TITLE OF INVENTION: POLYNUCLEOTIDE SEQUENCES AND CORRESPONDING ENCODED POLYPEPTIDES
; TITLE OF INVENTION: OF PARTICULAR SECRETED AND MEMBRANE-BOUND PROTEINS OVEREXPRESSED
; TITLE OF INVENTION: IN CERTAIN CANCERS
; FILE REFERENCE: Mergen - 0010B
; CURRENT APPLICATION NUMBER: US/10/775,920
; PRIOR FILING DATE: 2004-02-10
; PRIOR APPLICATION NUMBER: US 60/447,900
; PRIOR FILING DATE: 2003-02-13
; NUMBER OF SEQ ID NOS: 385
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 188
; LENGTH: 3171
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-775-920-188
Alignment Scores:
Pred. No.: 0 Length: 3171
Score: 4369.00 Matches: 829
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 17 Gaps: 0
US-09-916-849A-1 (1-829) x US-10-775-920-188 (1-3171)
QY 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuLeuLeuValCysTrpLeuGln 20
DB 54 ATGGGCTCCCTCGTGAGACCTCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 113
QY 21 CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
DB 114 TGGCGGCTCCAGCCCTGCGCGGCTCTTTCAGGAGGCTGAAGTGAAGTGAAGTGAAGTGAAG 173
QY 41 GlyGlyAlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
DB 174 GGAGGCGCGAGCAGGAGCG 233
QY 61 GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr 80
DB 234 CAAGACCCAGCTCTGTTTATGACTGATATGATGATGATGATGATGATGATGATGATGATGAT 293
QY 81 ValGlnGluArgArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
DB 294 GTCCAGCAAGAAGGTCTACTGAAGGAAGGAATCCATTGAAGATCTTCCCATCCCAAGCT 353
QY 101 IleLeuArgArgHisLysArgAspTrpValAlaProIleSerValProGluAsnGly 120

Db 354 ATCTACGACACACAGAGATTGGGTGGTTCCTCAATATCTGTCCCTGAAAAATGGC 413
Qy 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgThrLys 140
Db 414 AAGGGTCCTTCCCCACAGAGACTGAATCAGCTCAAGTCTAATAAAGATAGAGACCAAG 473
Qy 141 IlePheTyrSerIleThrGlyProGluValAlaAspSerProGluGluValPheAlaVal 160
Db 474 ATTTTCTACACATCACGGGCGCGGGGACAGACCCCTCGAGGGTGCTTCGCTGTA 533
Qy 161 GluLysGluThrGlyTyrLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys 180
Db 534 GAGAAGGAGACAGGCTGGTGTGTGAATAAGCACTGGACCGGAGGAGATTGCCAAG 593
Qy 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
Db 594 TATGAGTCTTTGGCCACGCTGTGTACAGAAATGGTGCCTCAGTGGAGGACCCCATGAC 653
Qy 201 IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
Db 654 ATCTCCATCATCGTGACGACACCAATGACACACAGCCCAAGTTTACCCAGGACACCTTC 713
Qy 221 ArgGlySerValLeuGluGluValLeuProGlyThrSerValMetGlnValThrAlaThr 240
Db 714 CGAGGGAGTGTCTTAGAGGGAGTCTTACAGGTACTTCTGTGATGCAGGTGACAGCCACA 773
Qy 241 AspGluAspAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln 260
Db 774 GATGAGGATGATGCATCTACCTCAATGGGTGGTGTGTACTCACTCATCCATAGCCAA 833
Qy 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
Db 834 GAACCAAGGACCCACACGACCTCATGTTTCAATTCACCGGAGCACAGGCACCATCAGC 893
Qy 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
Db 894 GTCACTCCAGTGCGCTGGACCGGAAAAGTCTCTGAGTACACACTGACCATCCAGGCC 953
Qy 301 ThrAspMetAspGlyAspGlySerThrThrAlaValAlaValAlaValGluIleLeuAsp 320
Db 954 ACAGACATGGATGGGACGGCTCCACCCACCGCAGTGGCAGTAGTGGAGATCCTTGAT 1013
Qy 321 AlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
Db 1014 GCCAATGACATGCTCCCATGTTTGGACCCCGAAGTACGAGGCCCATGTGCGCTGAGAT 1073
Qy 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
Db 1074 CCAGTGGGCCATGAGGTGCAGAGCTGACGGTCACTGATCTGACGCGCCCCCACTACCA 1133
Qy 361 AlaTrpArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
Db 1134 CGGTGGCGTGCACCTACCTTATCATGGCGGTGACGACGGGACCAATTTTACCATCACC 1193
Qy 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
Db 1194 ACCACCCCTGAGACACACAGGATCTCTGACACACCGAAGGGTTTGGATTTTGGAGCC 1253
Qy 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db 1254 AAAAAACGACACACCTGTGAGTGAAGTACCAACGAGGCCCTTTTGTGCTGAGACTC 1313
Qy 421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
Db 1314 CCAACCTCCACAGCCACCATAGTGGTCCAGTGGAGGATGTGAATGAGGACCTGTGTTT 1373
Qy 441 ValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
Db 1374 GTCCCAACCTCCAAAGTCTGTGAGTCTCAGAGGGGATCCCCACTGGGGAGCCTGTGTGT 1433
Qy 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db 1434 GTCTACACTGCAGAAAGCCCTGACAGAGAAATCAAAAGATCAGCTACCGCATCTCTGAGA 1493

Qy 481 AspProAlaGlyTyrTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db 1494 GACCCAGCAGGGTGGTAGCCATGGACCCAGACAGTGGGCGAGTCAACAGCTGTGGCACC 1553
Qy 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db 1554 CTCGACCGTGGAGATGAGCAGTTTGTGAGAAACACATCATTAAGATCATGTCTTGGCC 1613
Qy 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuLeuAsp 540
Db 1614 ATGGCAATGGAAGCCCTCCACCACTGGACCGGAACCTTCTGCTAAACATGATTGAT 1673
Qy 541 ValAsnAsnHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 1674 GTCAACGACCATGGCCCGCTGAGCCCTGAGCTCAGATCACCATCTGCAACCAAGCCCT 1733
Qy 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
Db 1734 GTGGCCACGCTGTGTAACATCACGGACAGGACCTGTCTCCACACACCTCCCTTCCAG 1793
Qy 581 AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGlyAsp 600
Db 1794 GCCACGCTCACAGATGACTCAGACATCTACTGACGGCAGAGGTCAACGAGAAAGGTGAC 1853
Qy 601 ThrValValLeuSerLeuLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
Db 1854 ACAGTGGTCTTCTCCCTGAAAGATTCTTGAAGCAGGATACATATGACGTGACCTTCT 1913
Qy 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
Db 1914 CTGCTCTGACATGGCAACAAAGAGAGCTGACGGTGTATCAGGGCCACTGTGTGCGACTGC 1973
Qy 641 HisGlyHisValGluThrCysProGlyProTyrLysGlyGlyPheIleLeuProValLeu 660
Db 1974 CATGGCCATGTGAAACCTGCCCCTGACCTCGAAGAGGAGTTTCATCTCCTCTGTGCTG 2033
Qy 661 GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysLys 680
Db 2034 GGGGTGTCTCTGCTGTCT 2093
Qy 681 ArgLysIleGlyGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
Db 2094 CGGAAGATCAAGAGAGCCCT 2153
Qy 701 TyrGlyGluGluGlyGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
Db 2154 TATGGCAAGAGAGGGGGTGGCGAGAGGACCAAGGACTATGACATCACCCAGCTCCACCA 2213
Qy 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleLeuPro 740
Db 2214 GGTCTGGAGGCCAGGCGGAGGTGTCTCTCGCAATGACGTGGCACCACCACTATCCCG 2273
Qy 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleLeuGlu 760
Db 2274 ACACCATGTACCGTCTTAGGCCAGCAACCCAGATGAAATCGGCACTTTTATAATTGAG 2333
Qy 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
Db 2334 AACCTGAAGGGCGCTAAACACAGACCCACAGCCCGCCCTACGACACCTCTTGTGTTC 2393
Qy 781 AspTyrGluGlySerGlySerAspAlaSerLeuSerSerLeuThrSerSerAlaSer 800
Db 2394 GACTATGAGGGCAGCGGTCCGACCGCGCTCCCTGAGCTCCCTCACCTCTCCGCCCTCC 2453
Qy 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTyrGlySerArgPheLysLysLeuAla 820
Db 2454 GACCAAGACCAAGATTACGATTATCTGAACGAGTGGGCGAGCGCTTCAAGAGGCTGGCA 2513
Qy 821 AspMetTyrGlyGlyGlyGluAspAsp 829
Db 2514 GACATGTACGGTGGCGGGGAGGACGAC 2540

RESULT 15

US-10-116-802-6
 ; Sequence 6, Application US/10116802
 ; Publication No. US20030065157A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Amy Lasek
 ; TITLE OF INVENTION: GENES EXPRESSED IN LUNG CANCER
 ; FILE REFERENCE: PA-0045 US
 ; CURRENT APPLICATION NUMBER: US/10/116.802
 ; CURRENT FILING DATE: 2002-04-04
 ; PRIOR APPLICATION NUMBER: 60/281,593
 ; PRIOR FILING DATE: 2001-04-04
 ; NUMBER OF SEQ ID NOS: 519
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 6
 ; LENGTH: 3777
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; OTHER INFORMATION: Incyte ID No: 331908.5
 US-10-116-802-6

Alignment Scores:

Pred. No.: 0 Length: 3777
 Score: 4369.00 Matches: 829
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 13 Gaps: 0

US-09-916-849A-1 (1-829) x US-10-116-802-6 (1-3777)

QY	201	IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe	220
DB	1216	ATCTCCATCATCGTACCGACGAGAAATGACCAAGGCCAAAGTTTACCAGGACACCTTC	12175
QY	221	ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr	240
DB	1276	CGAGGGAGTGTCTTAGAGGGAGTCTACAGGTACTTCTGTGATGCAGGTGACAGCCACA	1335
QY	241	AspGluAspAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln	260
DB	1336	GATGAGGATGATGCCATCTACACCTTACAAATGGGTGGTTGCTTACTCATCTAGCCAA	1395
QY	261	GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer	280
DB	1396	GAACCAAGGACCCACAGCTCATGTTCACCATTCACCGAGGACACAGGACCATCAGC	1455
QY	281	ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla	300
DB	1456	GTCACTCTCCAGTGGCTGGACCGGGAATAAGTCCCTGAGTACACACTGACCATCCAGGC	1515
QY	301	ThrAspMetAspGlyAspGlySerThrThrAlaValAlaValAlaValGluIleLeuAsp	320
DB	1516	ACAGACATGGATGGGGACGGCTCCACACCGGAGTGGCAGTAGTGGAGATCTTGAT	1575
QY	321	AlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn	340
DB	1576	GCCATGACAAATGCTCCCATGTTTGACCCCGAGAGTACGAGGCCCATGTGCTTGAGAA	1635
QY	341	AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro	360
DB	1636	GCAGTGGGCCATGAGGTGACAGGCTGACGTCTGATCTGGAGCGCCCACTACCA	1695
QY	361	AlaTyrArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr	380
DB	1696	GGCTGGGTGCCACCTACTTATCATGGCGGTGACGCGGGACCACTTTTACCATCACC	1755
QY	381	ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla	400
DB	1756	ACCCACCTCGAGAGCAACAGGGCATCTGACCAACAGGAAGGGTTTGGATTTTGAGGCC	1815
QY	401	LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu	420
DB	1816	AAAAACCCAGCACACCTGTACCTTGAAGTGACCAACAGAGGCCCTTTTGTCTGAGGCTC	1875
QY	421	ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe	440
DB	1876	CCAACTCCACAGCCACCATAGTGTGCTCCAGTGGAGGATGTAATGAGGCACCTGTGTTT	1935
QY	441	ValProProSerLysValValGluValGluGlyIleProThrGlyGluProValCys	460
DB	1936	GTCCCAACCTCCAAAGTCTGTAGGTCCAGAGGGCATCCCACTGGGGAGCCCTGTGTGT	1995
QY	461	ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg	480
DB	1996	GTCTACACTGCAGAGACCTTGACAGGAGAAATCAAAAGATCAGCTACCGCATCTGAGA	2055
QY	481	AspProAlaGlyTyrPheAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr	500
DB	2056	GAACCCAGCAGGGTGGCTAGCCATGGACCCAGACAGTGGGCGAGGTACAGCTGTGGGCAC	2115
QY	501	LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla	520
DB	2116	CTCGACCGTGAAGTACGAGTTTGTGAGGAACAACATCTATGAAGTCTATGCTCTGGCC	2175
QY	521	MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrIleLeuAsp	540
DB	2176	ATGGCAATGGAAGCCCTCCACCACTGGCAGGGGAACCTTCTGCTAACACTGATTGAT	2235
QY	541	ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro	560
DB	2236	GTCAACAGCACATGGCCCAAGTCCCTGAGGCCCTGAGTACCACTCTGCAACCAAGCCCT	2295
QY	561	ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln	580

QY	1	MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuGlnValCysTrpLeuGln	20
DB	616	ATGGGGTCCCTCGTGGACCTCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	675
QY	21	CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla	40
DB	676	TGGCGGGCTCCGAGCGCTGCGGGCGGTCTTCAAGGAGGCTGAGTGACCTTGGAGCG	735
QY	41	GlyGlyAlaGluGlnGluProGlyGlyAlaLeuGlyLysValPheMetGlyCysProGly	60
DB	736	GGAGCGCGGAGCAGGAGCCCGCCAGCGGTGGGGAAAGTATTTCATGGGCTGCCCTGGG	795
QY	61	GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr	80
DB	796	CAAGAGCAGCTCTGTTTGTAGCACGTGATAATGATGACTTCACTGTGCGGATGGCGAGA	855
QY	81	ValGlnGluArgArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg	100
DB	856	GTCCAGGAAAGAGGTCACTGAAGGAAGGAATCCATTGAAGATCTTCCATCCAAACGT	915
QY	101	IleLeuArgArgHisLysArgAspTrpValValAlaProIleSerValProGluAsnGly	120
DB	916	ATCTTACGAAGACACAAGAGAGATGGGTGGTCTCCAAATATCTGCTGGAATAATGGC	975
QY	121	LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys	140
DB	976	AAGGTCTCTTCCCCCAGAGACTGAATCAGCTCAAGTCTAATTAAGATAGACACCAAG	1035
QY	141	IlePheTyrSerIleThrGlyProGlyAlaAspSerProGluGlyValPheAlaVal	160
DB	1036	ATTTTCTACAGCATCACCGGGCGCGGGGACAGAGCCCTCGAGGGTGTCTTCGCTGTA	1095
QY	161	GluLysGluThrGlyTyrLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys	180
DB	1096	GAGAAAGAGACAGGCTGTGTTGTTGNATGAAGCACTGGACCGGGAGGAGATTGCCAAG	1155
QY	181	TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn	200
DB	1156	TATGAGCTTTTGGCCACGCTGTGTCTCAGAGAATGGTCTCAGTGGAGGAGCCCATGAAC	1215

Db	2296	GTGGCCACGTCGTGAACATCAGGACAAGACCTGCTCCACACCTCCCTTTCCAG	2355
Qy	581	AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGluGlyAsp	600
Db	2356	GCCAGCTCACAGATGACTCAGACATCTACTGGACGGCAGAGGTCAACGAGGAGGTGAC	2415
Qy	601	ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer	620
Db	2416	ACAGTGGTCTTGTCCCTCAAGAAGTCTCTGAAGCAGGATACATATGACGTGCACCTTCT	2475
Qy	621	LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys	640
Db	2476	CTGTCTGACCATGGCAACAAGAGCAGCTGACGGTGATCAGGGCCACTGTGTGCACTGC	2535
Qy	641	HisGlyHisValGluThrCysProGlyProTrpLysGlyGlyPheIleLeuProValLeu	660
Db	2536	CATGGCCATGTCGAAACCTGCTCCGACCTCGAAGGAGGAGGTTCATCCTCCTGTGCTG	2595
Qy	661	GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysLys	680
Db	2596	GGGGCTGTCCCTGGCTGCTGCTTCTCTCTGCTGCTGCTGCTTTTGGTGAGAAAG	2655
Qy	681	ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr	700
Db	2656	CGGAAGATCAAGGAGCCCTCTACTCCAGAGATGACACCCGTGACAAGCTCTTCTAC	2715
Qy	701	TyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArg	720
Db	2716	TATGGCAAGAGGGGGTGGGAGAGAGACAGGACTATGACATCACCAGCTCCACCGA	2775
Qy	721	GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro	740
Db	2776	GGTCTGGAGGCCAGCGCGAGGTGGTCTCCGCANTGACGTGGCACCACCACCATCCCG	2835
Qy	741	ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu	760
Db	2836	ACACCCATGTACCGTCTTAGCCAGCCCAACCCAGATGAAATCGGCAACTTTATAATTGAG	2895
Qy	761	AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe	780
Db	2896	AACCTGAAGCGGGTTACACAGACCCACAGCCCGCCCTACGACACCCCTCTGGTGTTC	2955
Qy	781	AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer	800
Db	2956	GACTATGAGGCGAGCGGCTCCGACGCGCGTCCCTGAGTCCCTCACCTCCTCCGCTCC	3015
Qy	801	AspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGlySerArgPheLysLysLeuAla	820
Db	3016	GACCAAGACCAAGATTACGATTATCTGAACGAGTGGGGCAGCCGCTTCAAGAAGCTGGCA	3075
Qy	821	AspMetTyrGlyGlyGluAspAsp	829
Db	3076	GACATGTACGTGGGGGAGGACGAC	3102

Search completed: September 23, 2004, 03:46:12
Job time : 1023 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 22:03:42 ; Search time 21 Seconds
(without alignments)
3797.275 Million cell updates/sec

Title: US-09-916-849A-1
Perfect score: 4369
Sequence: 1 MGLPRGPLASLLLVQVWLQ.....NEWGRFVKLADMYGGEDD 829

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4369	100.0	829	1 IJHUCP	cadherin 3 precurs
2	3555.5	81.4	822	1 IJMSCF	P-cadherin precurs
3	2573.5	58.9	732	1 IJCHCB	B-cadherin precurs
4	2420	55.4	882	1 IJHUCE	cadherin 1 precurs
5	2388	54.7	884	1 IJMSCE	E-cadherin precurs
6	2379	54.5	884	2 S34438	uvomorulin - mouse
7	2375	54.4	895	1 IJXLCF	EP-cadherin precurs
8	2371	54.3	905	2 S43064	cadherin - African
9	2323	53.2	491	1 IJBOCP	P-cadherin - bovin
10	2274	52.0	887	1 IJCHCL	P-cadherin precurs
11	2192.5	50.2	871	2 S47518	cadherin - African
12	1707	39.1	905	1 IJHUCN	cadherin 2 precurs
13	1704.5	39.0	912	1 IJHUCN	N-cadherin precurs
14	1697.5	38.9	906	1 IJMSCN	N-cadherin precurs
15	1694	38.8	913	1 IJCHCR	R-cadherin precurs
16	1691	38.7	877	1 IJBOCN	N-cadherin precurs
17	1655	37.9	913	1 A47543	R-cadherin precurs
18	1621.5	37.1	916	2 C38992	cadherin 4 precurs
19	1621	37.1	783	2 I50116	N-cadherin precurs
20	1616.5	37.0	905	1 IJXLC2	N-cadherin 2 precu
21	1609	36.8	905	1 IJXLC1	N-cadherin 1 precu
22	1321.5	30.2	814	2 G02878	cadherin-15 precu
23	1302.5	29.8	730	1 IJMSCM	M-cadherin - mouse
24	1021.5	23.4	713	2 B38992	cadherin 13 precu
25	1007.5	23.1	785	2 I50180	cadherin-7 - chick
26	990	22.7	790	2 I50178	cadherin-6B - chic
27	982	22.5	790	2 I37016	cadherin-6 - human
28	976.5	22.4	790	2 G02678	cadherin-14 - huma
29	976	22.3	712	1 IJMSCT	T-cadherin precurs

ALIGNMENTS

RESULT 1

IJHUCP
cadherin 3 precursor - human
N;Alternate names: P-cadherin; placental cadherin
C;Species: Homo sapiens (man)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C;Accession: A33659
R;Shimoyama, Y.; Yoshida, T.; Terada, M.; Shimosato, Y.; Abe, O.; Hirahashi, S.
J. Cell Biol. 109, 1787-1794, 1989
A;Title: Molecular cloning of a human Ca(2+)-dependent cell-cell adhesion molecule homolog
A;Reference number: A33659; MUID:90009051; PMID:2793940
A;Accession: A33659
A;Molecule type: mRNA
A;Residues: 1-829 <SHI>
A;Cross-references: GB:X63629; NID:G35322; PIDN:CAA45177.1; PID:G35323
C;Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought to
C;Genetics:
A;Gene: GDB:CDH3
A;Cross-references: GDB:132860; OMIM:114021
A;Map position: 16q24.1-16qter
C;Superfamily: cadherin; cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane prot
F;1-26/Domain: signal sequence #status predicted <SIG>
F;27-107/Domain: propeptide #status predicted <PRO>
F;108-829/Product: P-cadherin #status predicted <MAT>
F;110-215/Domain: extracellular #status predicted <EXT>
F;185-190/Region: cadherin repeat homology <CR1>
F;218-328/Domain: cadherin binding #status predicted
F;331-440/Domain: cadherin repeat homology <CR2>
F;441-548/Domain: cadherin repeat homology <CR3>
F;549-652/Domain: cadherin repeat homology <CR4>
F;653-677/Domain: transmembrane #status predicted <TMM>
F;678-829/Domain: intracellular #status predicted <INT>
F;785-800/Region: serine-rich
F;200,566/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 4369; DB 1; Length 829;
Best Local Similarity 100.0%; Pred. No. 8.4e-271;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MGLPRGPLASLLLVQVWLQVCAASPCRAVFEASVTLEAGCAQEPGQALCKVFMGCG 60
Db	1	MGLPRGPLASLLLVQVWLQVCAASPCRAVFEASVTLEAGCAQEPGQALCKVFMGCG 60
QY	61	QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDWVAPISVPENG 120
Db	61	QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDWVAPISVPENG 120
QY	121	KGPFQRLNQLSKNRDRTKIFYSITGCASSPPGFAVEKETGWLILNKLDPDEETAK 180
Db	121	KGPFQRLNQLSKNRDRTKIFYSITGCASSPPGFAVEKETGWLILNKLDPDEETAK 180

```

181 YELRGHVAENGASVEDPMNLSIIIVTDNDHKPKFTQDTRGSLVGLVPGTSMQVAT 240
181 YELRGHVAENGASVEDPMNLSIIIVTDNDHKPKFTQDTRGSLVGLVPGTSMQVAT 240
241 DEDDAIYTYNGVWVAYSIIHQSPKDPHDLMTIHRSTGTISVISSGLDREKVPYTLTIOA 300
241 DEDDAIYTYNGVWVAYSIIHQSPKDPHDLMTIHRSTGTISVISSGLDREKVPYTLTIOA 300
301 TDMGDSGTTTAVAVVLLDANDNAPMPDPQKYEAHVPEAVGHEVORLTVDLADNSP 360
301 TDMGDSGTTTAVAVVLLDANDNAPMPDPQKYEAHVPEAVGHEVORLTVDLADNSP 360
361 AWRATYILMGDDGDDGHFTITTHPSNQGLITRKGLOFEAKNQHTLYVEVNEAPFVILK 420
361 AWRATYILMGDDGDDGHFTITTHPSNQGLITRKGLOFEAKNQHTLYVEVNEAPFVILK 420
421 PTSTATIIVHVEDVNEAPVFPVPPSKVVEVQGIPTGEPVCYVTAEDDPKENOKISYRILR 480
421 PTSTATIIVHVEDVNEAPVFPVPPSKVVEVQGIPTGEPVCYVTAEDDPKENOKISYRILR 480
481 DPAGWLAMPDPSGGVAVTAVGLDREDEQFVRNNIYEVVVLAMDNQSPPTTGTGTLTLLD 540
481 DPAGWLAMPDPSGGVAVTAVGLDREDEQFVRNNIYEVVVLAMDNQSPPTTGTGTLTLLD 540
541 VNDHGVPEPEPQITICNQSPVRHVNITDKDLSPTSPFOQLTDDSDIYWTAEVNEEGD 600
541 VNDHGVPEPEPQITICNQSPVRHVNITDKDLSPTSPFOQLTDDSDIYWTAEVNEEGD 600
601 TVVLSLKKFLKQDVTYVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWKGFIPLVL 660
601 TVVLSLKKFLKQDVTYVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWKGFIPLVL 660
661 GAVLALLFLLVLLVLLVLRKKRKEPILLPPDDTRDNVYFYGEGGEEEDQDYDITQJHR 720
661 GAVLALLFLLVLLVLLVLRKKRKEPILLPPDDTRDNVYFYGEGGEEEDQDYDITQJHR 720
721 GLEAPPEVLRNDVAPTITPMYRPRPANDPEIGNFIENLKAANTDPTAPPYDTLLVF 780
721 GLEAPPEVLRNDVAPTITPMYRPRPANDPEIGNFIENLKAANTDPTAPPYDTLLVF 780
781 DYEGSGSDAASLSLTSASDQDQDYDYLNEGSRFKKLADMYCGGEDD 829
781 DYEGSGSDAASLSLTSASDQDQDYDYLNEGSRFKKLADMYCGGEDD 829

RESULT 2
IUMSCP
p-cadherin precursor - mouse
C:Species: Mus musculus (house mouse)
C>Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 07-May-1999
C:Accession: S03163; S34458
R:Nose, A.; Nagafuchi, A.; Takeichi, M.
EMBO J. 6, 3655-3661, 1987
A:Title: Isolation of placental cadherin cDNA: identification of a novel gene family of
A:Reference number: S03163; MUID:86111554; PMID:3428270
A:Accession: S03163
A:Molecule type: mRNA
A:Residues: 1-822 <NOS>
A:Cross-references: EMBL:X06340
R:Faraldo, M.L.M.; Cano, A.
J. Mol. Biol. 231, 935-941, 1993
A:Title: The 5' flanking sequences of the mouse p-cadherin gene. Homologies to 5' sequen
A:Reference number: S34458; MUID:93294853; PMID:8515462
A:Accession: S34458
A:Molecule type: DNA
A:Residues: 1-55 <FAR>
A:Cross-references: EMBL:X68057
C:Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought b
C:Genetics:
A:Introns: 16/3
C:Superfamily: cadherin; cadherin repeat homology
C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; placenta; transme

```

```

F:1-27/Domain: signal sequence #status predicted <SIG>
F:28-99/Domain: propeptide #status predicted <PRO>
F:100-822/Product: P-cadherin #status predicted <MAT>
F:100-645/Domain: extracellular #status predicted <EXT>
F:102-207/Domain: cadherin repeat homology <CR1>
F:210-320/Domain: cadherin repeat homology <CR2>
F:323-432/Domain: cadherin repeat homology <CR3>
F:433-540/Domain: cadherin repeat homology <CR4>
F:541-645/Domain: cadherin repeat homology <CR5>
F:646-670/Domain: transmembrane #status predicted <TM>
F:671-822/Domain: intracellular #status predicted <INT>
F:778-793/Region: serine-rich
F:192,558/Binding site: carboxydrate (Asn) (covalent) #status predicted

Query Match      81.4%; Score 3555.5; DB 1; Length 822;
Best Local Similarity 81.6%; Pred. No. 6.8e-219;
Matches 679; Conservative 57; Mismatches 83; Indels 13; Gaps 5;

QY 1 MGLPRGPIA-SLLILQVCMLOCAASEPCRAVF-REAEVTLAAGAEQEPGQALGKVFNGC 58
DB 1 MELLSPHAFLLLLQVCMLSVSEPYRAGFIGEAGVTLEVEGTDLEPSQVLGKVALAG 60
QY 59 PQCEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDXVAVAPISYE 118
DB 61 QGMHHA-----DNGDIIMLTRGTVOGKDAHS-----PPTRILRRKREWMPPIFVPE 110
QY 119 NGKGFPPQRLNOLKNDKRDTKIFYSITGPGADSPPEGVFAVEKETHGLLLINKPLDREEL 178
DB 111 NGKGFPPQRLNOLKNDKRGTKIFYSITGPGADSPPEGVFTIEKESGMLLLHMLDREKI 170
QY 179 AKVELFCHAVSNGASVEDPMNLSIIIVTDNDHKPKFTQDTRGSLVGLVPGTSMQVOT 238
DB 171 VKYELFCHAVSNGASVEDPMNLSIIIVTDNDHKPKFTQDTRGSLVGLVPGTSMQVOT 230
QY 239 ATDEDDAIYTYNGVWVAYSIIHQSPKDPHDLMTIHRSTGTISVISSGLDREKVPYTLT 298
DB 231 ATDEDDAVNTYNGVWVAYSIIHQSPKDPHDLMTIHRSTGTISVISSGLDREKVPYTLT 290
QY 299 QATMDGDSGTTTAVAVVLLDANDNAPMPDPQKYEAHVPEAVGHEVORLTVDLADPN 358
DB 291 QATMDGDSGTTTAVAVVLLDANDNAPMPDPQKYEAHVPEAVGHEVORLTVDLADPN 350
QY 359 SPAPRATYILMGDDGDDGHFTITTHPSNQGLITRKGLOFEAKNQHTLYVEVNEAPFVL 418
DB 351 WPAPRATYIIVGDDGDDGHFTITTHPSNQGLITRKGLOFEAKNQHTLYVEVNEAPFVL 410
QY 419 KLPTSTATIIVHVEDVNEAPVFPVPPSKVVEVQGIPTGEPVCYVTAEDDPKENOKISYRI 478
DB 411 KLPTATATVTVHVKDNEAPVFPVPPSKVVEVQGIPTGEPVCYVTAEDDPKENOKISYRI 470
QY 479 LRDPAGWLAMPDPSGGVAVTAVGLDREDEQFVRNNIYEVVVLAMDNQSPPTTGTGTLTLL 538
DB 471 SRDPANLAVDPDSGQITPAAGILDREDEQFVRNNIYEVVVLAMDNQSPPTTGTGTLTLL 530
QY 539 IDNVHGPVPEPQITICNQSPVRHVNITDKDLSPTSPFOQLTDDSDIYWTAEVNEE 598
DB 531 TDINDHGPVPEPQITICNQSPVRHVNITDKDLSPTSPFOQLTDDSDIYWTAEVNEE 590
QY 599 GDTVVLKLLKQDVTYVHLSLSDHGNKEQLTVIRATVCDCHGHV-ETCPQPKWKGFIPL 657
DB 591 GDTVALSLKFKLQDVTYVHLSLSDHGNKEQLTVIRATVCDCHGHV-ETCPQPKWKGFIPL 650
QY 658 FVLGAVLALLFLLVLLVLRKKRKEPILLPPDDTRDNVYFYGEGGEEEDQDYDITQ 717
DB 651 FVLGAVLALLFLLVLLVLRKKRKEPILLPPDDTRDNVYFYGEGGEEEDQDYDITQ 710
QY 718 LHRGLEARPEVLRNDVAPTITPMYRPRPANDPEIGNFIENLKAANTDPTAPPYDTLL 777
DB 711 LHRGLEARPEVLRNDVAPTITPMYRPRPANDPEIGNFIENLKAANTDPTAPPYDTLL 770
QY 778 LVFDYEGSGSDAASLSLTSASDQDQDYDYLNEGSRFKKLADMYCGGEDD 829
DB 771 LVFDYEGSGSDAASLSLTSASDQDQDYDYLNEGSRFKKLADMYCGGEDD 822

```

```

RESULT 3
LCHCS
B-cadherin precursor - chicken (fragment)
N:Alternate names: K-CAM protein
C:Species: Gallus gallus (Chicken)
C>Date: 30-Jun-1993 #sequence revision 30-Jun-1993 #text_change 22-Jun-1999
C:Accession: A41634; A38715; S16160
R:Sorkin, B.C.; Gallin, W.J.; Edelman, G.M.; Cunningham, B.A.
Proc. Natl. Acad. Sci. U.S.A. 88, 11545-11549, 1991
A:Title: Genes for two calcium-dependent cell adhesion molecules have similar structures
A:Reference number: A41634; MUID:92107987; PMID:1763068
A:Accession: A41634
A:Molecule type: DNA
A:Residues: 1-732 <SOR>
A:Cross-references: GB:M81894; NID:g212226; PIDN:AAA48929.1; PID:g212227
R:Napolitano, E.W.; Venstrom, K.; Wheeler, E.F.; Reichardt, L.F.
J. Cell Biol. 113, 893-905, 1991
A:Title: Molecular cloning and characterization of B-cadherin, a novel chick cadherin.
A:Reference number: A38715; MUID:91225083; PMID:2026653
A:Accession: A38715
A:Molecule type: mRNA
A:Residues: 7-413, 'V', 415-732 <NAP>
A:Cross-references: GB:X58518; NID:g63113; PIDN:CAA41408.1; PID:g63114
C:Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought to
C:Genetics:
A:Gene: K-CAM
A:Introns: 29/3; 81/3; 130/1; 188/3; 231/3; 293/3; 375/2; 423/1; 498/1; 571/1; 614/3; 666/3
C:Superfamily: cadherin; cadherin repeat homology
C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane protein
F:1-6/Domain: propeptide (fragment) #status predicted <PRO>
F:6-554/Domain: extracellular #status predicted <EXT>
F:7-732/Product: B-cadherin #status predicted <NAT>
F:9-114/Domain: cadherin repeat homology <CR1>
F:84-89/Region: cadherin binding #status predicted
F:117-227/Domain: cadherin repeat homology <CR2>
F:230-339/Domain: cadherin repeat homology <CR3>
F:340-447/Domain: cadherin repeat homology <CR4>
F:448-552/Domain: cadherin repeat homology <CR5>
F:555-580/Domain: transmembrane #status predicted <TM>
F:581-732/Domain: intracellular #status predicted <INT>
F:689-702/Region: serine-rich
F:137,410/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match      58.9%; Score 2573.5; DB 1; Length 732;
Best Local Similarity 66.2%; Pred. No. 2.7e-156;
Matches 485; Conservative 103; Mismatches 138; Indels 7; Gaps 5;

Qy 102 LRHKRDWVAPISVPENKGPFPQRLNQLKSNKRDTKIFYSITGFGADSPPEGVFAVE 161
Db 1 LRQKRDWVAPPIKVPENERGPPKNLVQIKSNRDEAKIFYSITGQGAADPEGIFTIE 60

Qy 162 KETGWLILNKLDRERIATKELFGHVAENGASVEDPMNISIIVTDNDHKPKFTQDIFR 221
Db 61 KETGWMKVTQPLDRHINKLTHLYSHVAENGKQVEBPEMIIIVTVDQNDNKQFTQVEFR 120

Qy 222 GSVLEGVLPQTSVMQVATDDEDAIYTYNGVAYSISHSQEPKDPHLMFTIHRSTGTSIV 281
Db 121 GSVPEGALPGTSVMRVNATDADDQVETNGVIAYSILSQEPREPHFMFTVARATGTLV 180

Qy 282 ISSGLDRKYPEYTLIIQATMDGDGSTTAVAVAILDANAPMDFDQKYEAVHPENA 341
Db 181 IASGLDRVRREYTLITMQAADLQDGLTTALAVISITVDNNAPEFDPKTYEAAVPENE 240

Qy 342 VGHEVORLITVDLDAPNSPAWRATYILMGDDGDHETITTHPESNOGIIITTRKGLDFSAK 401
Db 241 AELEVARLATLTDLDEPHTPAWRAVYISVRGNEGAFTITTDPASNEGVLRTAKGLDYEA 300

Qy 402 NQHTLVVEVNEAPFVLKLPSTATIVHVEDVNEAPFVFPKSVKVEOGIPTGEPCVCV 461
Db 301 RQFVLHVAVNEAPFAIKLPTATVMVSDVNEAPFVDFPRLAQVDFPVLQGLAS 360

```

```

Qy 462 YTAEDPDK-ENQKISYRILRDPAGWLANDPDSGGQVTAAGTLDREDQFVRNNIYEVNMLA 520
      |||:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:
Db 361 YTAQDPDRAQQRIKYVNGMSDPAGWLAHPENGLITAREQLDRE-SPTTKNSTYMAVLLA 419
      |||:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:
Qy 521 MDNGSPPTGTGTLTLTLIDVNDHGVFVPEPRQITICNQSPVRHVLNITDKLSPHTSPFQ 580
      |||:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:
Db 420 VDDGLPPATGTGTLTLTLVNDHGVPEPRDIVICNRSPVQVLTITDRDLPEPTGPR 479
      |||:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:
Qy 581 AQLTDDSDIYTAENVNEGDITVLSLKKFLKQDTYDVHLSLSHGNKKEQLTVIRATVDCD 640
      |||:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:
Db 480 AELSHSGSDS WAVEVNGGSDTVALWLTEPLEQNLSVYLEFDRQGDQDVTVIRAQVDCD 539
      |||:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:
Qy 641 HGHVETC---PQPMKG-GFILLPVLCAVALLLELLVLLLVKKRKIKPELLLPEDDTRD 696
      |||:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:
Db 540 QGRVESCQAQKPRVDTGVPIVLAVLGAVLALLVLLVLLLVRRKVKVKEPLLIPEDDTRD 599
      |||:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:
Qy 697 NVFYGBGGGEEBDDYDITQLHRLGARPEVWLNRNDVAPTITPTMYRPRPANPEIGN 756
      |||:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:
Db 600 NIFYGBGGGEEQDYDLSQLHRLGARPEVI-RNDVAPFLMAAPQYRPRPANPEIGN 658
      |||:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:
Qy 757 FIENLKAANDTPATPPYDITLLVFFYEGSGSAASSLTSSASDQDQDYDYLNEGSRF 816
      |||:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:
Db 659 FIDENLKAADTTPATPPYDLSLLVFFYEGGSEATSLSSLNSSASDQDQDYDYLNEGNRF 718
      |||:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:
Qy 817 KKLADMTYGGGEBD 829
      |||:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:
Db 719 KKLAELYGGGEDE 731
      |||:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:

RESULT 4
LHUCE
catherin 1 precursor [validated] - human
N/Alternate names: ARC-1; cell CAM 120/80; E-cadherin; epithelial cadherin; L-CAM; uvomor-
C/Species: Homo sapiens (man)
C/Date: 30-Jun-1993 #sequence revision 30-Jun-1993 #text change 08-Dec-2000
C/Accession: S31430; S05475; S31460; S06716; A5T171; J02230; I52294; I52704; S251
R/Bussmakers, M.J.G.; van Bokhoven, A.; Mees, S.G.M.; Kemler, R.; Schalken, J.A.
Mol. Biol. Rep. 17, 123-128, 1993
A/Title: Molecular cloning and characterization of the human E-cadherin cDNA.
A/Reference number: S37654; MUID:93211394; PMID:8459805
A/Accession: S37654
A/Molecule type: mRNA
A/Residues: 1-882 <BUS>
A/Cross-references: EMBL:Z13009; NID:g31072; PIDN:CAA78353.1; PID:g31073
R/Kalker, W.; Warda, A.; Oda, T.; Hirohashi, S.; Kemler, R.; Birchmeier, W.
submitted to the EMBL Data Library, December 1992
A/Description: Sequence of human E-cadherin cDNA.
A/Reference number: S31430
A/Accession: S31430
A/Molecule type: mRNA
A/Residues: 1-542, 'F', 544-882 <KEL>
A/Cross-references: EMBL:Z18923; NID:g31074; PIDN:CAA79356.1; PID:g31075
R/Mansouri, A.; Spurr, N.; Goodfellow, P.N.; Kemler, R.
Differentiation 38, 67-71, 1998
A/Title: Characterization and chromosomal localization of the gene encoding the human ceJ
A/Reference number: S05475; MUID:89031725; PMID:3263290

```

A:Molecule type: protein
A:Residues: 157-162, 'V', 164-179 <WHE>
R:Brix, G.; Saes, K.; van Hengel, J.; Molemans, F.; Bussemakers, M.J.G.; van Bokhoven, Genomics 26, 281-289, 1995
A:Title: Cloning and characterization of the human invasion suppressor gene E-cadherin
A:Reference number: A57171; MUID:95324920; PMID:7601454
A:Accession: A57171
A:Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-30,32-882 <BER>
A:Cross-references: GB:L34784
R:Rimm, D.L.; Morrow, J.S.
Biochem. Biophys. Res. Commun. 200, 1754-1761, 1994
A:Title: Molecular cloning of human E-cadherin suggests a novel subdivision of the cadherin
A:Reference number: JC2230; MUID:94242050; PMID:8185635
A:Accession: JC2230
A:Molecule type: mRNA
A:Residues: 1-9, 'G', 11-15, 'RSPGQSRSPPCLTRELHVHGAPAPKEPR', 52-67, 'I', 69, 'LTPIP', 76-94
A:Cross-references: GB:L08599; NID:9340184; PID:AAA61259.1; PID:9340185
A:Note: the majority of differences between this and other reports represent apparent frameshifts
A:Note: the authors translated the codon CCG for residue 868 as Arg
R:Bussemakers, M.J.G.; Girolidi, L.A.; van Bokhoven, A.; Schalken, J.A.
Biochem. Biophys. Res. Commun. 203, 1284-1290, 1994
A:Title: Transcriptional regulation of the human E-cadherin gene in human prostate cancer
A:Reference number: 152294; MUID:94380041; PMID:8093045
A:Accession: 152294
A:Status: translation not shown; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-16 <RES>
A:Cross-references: GB:L34545; NID:G509604; PID:AAA21764.1; PID:G509605
R:Becker, K.F.; Atkinson, M.J.; Reich, U.; Becker, I.; Nekarda, H.; Siewert, J.R.; Hofler, Cancer Res. 54, 3845-3852, 1994
A:Title: E-cadherin gene mutations provide clues to diffuse type gastric carcinomas.
A:Reference number: 152704; MUID:94308394; PMID:8033105
A:Accession: 152704
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 33-476 <RE2>
A:Cross-references: GB:S72492; NID:G632756; PIDN:AA14108.1; PID:G4261808
A:Comment: Cadherins mediate calcium-dependent intercellular adhesion and are thought to be involved in many cellular processes
C:Genetics:
A:Gene: GDB:CDH1; UVO
A:Cross-references: GDB:120484; OMIM:192090
A:Map position: 16q22.1-16q22.1
A:Introns: 379/3; 440/3
C:Superfamily: cadherin; cadherin repeat homology
C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane protein
F:1-27/Domain: signal sequence #status predicted <SIG>
F:28-154/Domain: propeptide #status predicted <PRO>
F:155-882/Product: E-cadherin #status experimental <MAT>
F:155-697/Domain: extracellular #status predicted <EXT>
F:157-262/Domain: cadherin repeat homology <CR1>
F:232-237/Region: cadherin binding #status predicted
F:265-375/Domain: cadherin repeat homology <CR2>
F:378-486/Domain: cadherin repeat homology <CR3>
F:487-595/Domain: cadherin repeat homology <CR4>
F:596-700/Domain: cadherin repeat homology <CR5>
F:698-733/Domain: transmembrane #status predicted <TM>
F:732-882/Domain: intracellular #status predicted <INT>
F:840-853/Region: serine-rich
F:876,558,570,622,637,849/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 55.4%; Score 2420; DB 1; Length 882;
Best Local Similarity 55.1%; Pred. No. 2.2e-146;
Matches 483; Conservative 112; Mismatches 222; Indels 60; Gaps 13;

QY 9 ASLLILLOV-CWLCQASPCRAVFEAEVTEAGCAQALQKV-FMCPG-QEPAL 65
DB 10 ALLLLQVSSWL-CQEPFCHGFDASFTVTPRRHLGRVLRVNFEDCTGRTAY 68
QY 66 FSTNDNDFTVRNGETVQRRSLKERNP----- 92
DB 69 FSLDT-RFKVGDGVTIKRPLRFENPQILFLVAVWDSTYKFSKVTLTNVGHHRPPP 127

QY 93 -----LKITP-SKRILRRHKDWDVAPISVPENKGPQPPQRLNQLKSKKDRDK 140
DB 128 HQASVSGTQAEALLTFPNSPGLRQKRDWIPISCPENEXGPPKMLVQIKSNKDEKG 187
QY 141 IFYSTGTGADSPGEGVAFVAVKGTGWLKLLKPLDREBIAKVELFCHAVSNGASVEDPMN 200
DB 188 VFYSTGTGADTPPVGVFIIFRETCGLKXTEPLDREIATYTLFSAVSSNGNAVEDPME 247
QY 201 ISIIIVTDQNDHKPKFTQTPFGSLEGLVLPQTSVMQVTADEDDAIITYNGVAVSHSQ 260
DB 248 ILITVTDQNDKPEFTQSVFKGSVMEGALPGTSVMEVTAATDADDDVNTYNAAIATYTL 307
QY 261 BPKDPLDMFTIHRSTGTISVSSGLDREKVEYTLTIQATDMDGSGTTFVAVAVWILD 320
DB 308 DPELFDKKNFTINRTGIVSVTTGLDRESPTTILVVOAADLQEGSLSTATAVITVD 367
QY 321 ANDNAPMFDPOKYEAHVPENAVGHEVQRLTVTDLDAPNSPAWATYILMGDDGDGHFTIT 380
DB 368 TNDNPPINFTYKGVPEANEANVTITLKVTDADAPNTPAWEAVYTL-NDGCGQFVVT 426
QY 381 TPESNOGILTRKGLDPEAKNQHTLYVETNEAFVFLKLTSTATIVVHVEDVNEAPVF 440
DB 427 TNFVNDGILTKAGLDPEAKQVILHVAVTNVVFFEVSLTSTATVTDVLDVNEAPIF 486
QY 441 VPESKVEVQSGIPTGEPVCVVTAEDPK-ENQKISYRILRDPAGWLAMPDPSQGVAVG 499
DB 487 VPPEKRVESDGFVQISYTAQEPDTEQKITYRIWEDTANWLEINPDTGAISTRA 546
QY 500 TLDREDEQFVNNIYVWVWVMDNGSPPTGTGILLTLIDVNDHGPVPEPQITICNQ 559
DB 547 ELDRDFFHVKNSTYALITDNGSPVATGTGILLTLSDWNDNAPTEPRTIFFCERN 606
QY 560 PVRHVLTITKDLSPHTSPFOQLTDDSDIYVTAEVNE-EGDTVVLSLKFKQDTYDVH 618
DB 607 PKQVINIIDADLPNTSFTTAEILTHGASANTWTQYNDPTQESILILPKMALEVGYDKIN 666
QY 619 LSLSDHGKNEQLTVIRATVCDCHGHVETC--PGPMKGGF----ILPVGLVALLLILL 672
DB 667 LKLMNDQNKQDQVTTDEVSVCDCEGAAGVCRKAQPVAGLQIPAILGILGGLALLILL 726
QY 673 LILLVKKKIKERILLDEDDTRDNVYVYEGEGGEDDYDIQHLRGLBARPEWLRN 732
DB 727 LLLFLRRRAVKEPILLPPEDDTRDNVYIYDEGGEGEDQDFLSQLRGLDAREVT-RN 785
QY 733 DVAPTIIPPMYRPRPANDPDEIGNFIENLKAANTDPTAPPYDTLLVPDYEGSGSDAASL 792
DB 786 DVAPILMSVPRYLPRPANDPDEIGNFIDENLKAADTPTAPPYDLSLLVPDYEGSGSEASL 845
QY 793 SSLTSASDQDDYDLNWSRPFKILADMYGGGDD 829
DB 846 SSLNSSESDQDDYDLNWSRPFKILADMYGGGDD 882

RESULT 5
LUMSC
E-cadherin precursor, epithelial - mouse
N;Alternate names: uvomorulin
C;Species: Mus musculus (house mouse)
C;Date: 30-Jun-1993 #sequence revision 30-Jun-1993 #text_change 22-Jun-1999
C;Accession: S04528; S03160; I49565; S48735
R;Nagafuchi, A.; Shirayoshi, Y.; Okazaki, K.; Yasuda, K.; Takeichi, M.
Nature 329, 341-343, 1987
A;Title: Transformation of cell adhesion properties by exogenously introduced E-cadherin
A;Reference number: S04528; MUID:87315445; PMID:3498123
A;Accession: S04528
A:Molecule type: mRNA
A;Residues: 1-412, 'V', 414-884 <NAG>
A;Cross-references: EMBL:X06115
R;Ringwald, M.; Schuh, R.; Vestweber, D.; Eistetter, H.; Lottspeich, F.; Engel, J.; Doel; EMBO J. 6, 3647-3653, 1987
A;Title: The structure of cell adhesion molecule uvomorulin. Insights into the molecular
A;Reference number: S03160; MUID:88111553; PMID:3501370

A:Accession: S03160
A:Molecule type: RNA
A:Residues: 157-884 <RIN>
A:Cross-references: EMBL:X06339
A:Note: part of this sequence, including the amino end of the mature protein, was confirmed by Behrens, J.; Loewrick, O.; Klein-Hitpaas, L.; Birchmeier, W.
Proc. Natl. Acad. Sci. U.S.A. 88, 11495-11499, 1991
A:Title: The E-cadherin promoter: Functional analysis of a G-C-rich region and an epithelial reference number: 149565; MUID:92107977; PMID:11763063
A:Accession: I49565
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-15 <RES>
A:Cross-references: GB:M81449; NID:G192325; PIDN:AAA37352.1; PID:G192326
R:Tong, K.I.; Yau, P.; Overduin, M.; Bagby, S.; Porumb, T.; Takeichi, M.; Ikura, M.
FEBS Lett 352, 318-322, 1994
A:Title: Purification and spectroscopic characterization of a recombinant amino-terminal E-cadherin extracellular domain: NID:G192325; PIDN:AAA37352.1; PID:G192326
A:Reference number: S48735; MUID:95010732; PMID:7925993
A:Accession: S48735
A:Status: preliminary
A:Molecule type: protein
A:Residues: 156-300 <TON>
C:Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought to be involved in many cellular processes.
C:Genetics:
A:Gene: E-cadherin
C:Superfamily: cadherin; cadherin repeat homology
C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane protein
F:1-27/Domain: signal sequence #status predicted <SIG>
F:28-156/Domain: propeptide #status predicted <PRO>
F:157-884/Product: E-cadherin, epithelial #status experimental <MAT>
F:157-699/Domain: extracellular #status predicted <EXT>
F:159-264/Domain: cadherin repeat homology <CR1>
F:234-239/Region: cadherin binding #status predicted
F:267-377/Domain: cadherin repeat homology <CR2>
F:380-488/Domain: cadherin repeat homology <CR3>
F:489-597/Domain: cadherin repeat homology <CR4>
F:598-702/Domain: cadherin repeat homology <CR5>
F:702-733/Domain: transmembrane #status predicted <TM>
F:734-884/Domain: intracellular #status predicted <INT>
F:842-855/Region: serine-rich
F:560,639/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 54.78; Score 2388; DB 1; Length 884;
Best Local Similarity 53.6; Pred. No. 2.4e-444; Indels 66; Gaps 13;
Matches 472; Conservative 123; Mismatches 220; Indels 66; Gaps 13;

QY 9 ASLLQLQV-CWLOCAASEP--CRAVFEAEVTLAAGAEQEPGALGV-FMGCPGQEP 64
DB 10 ALLLLQLQVSSWL-COELEPSCSPGFSSEVYTFPVEHLERGHVLRVFECTGRPT 68
QY 65 LFSTNDNDFTVRNGETVOERSLK----- 88
DB 69 AFPSDSRFKVTADGTTITVKRHLKHLKLETSFLVRARDSSHRELSTKVTLSKMGHHHRH 128
QY 89 -----ERNP-LKIPSKRI-LRRHKRDWVAPISVPENKGGPPQRLNOLSKNDRDT 139
DB 129 HHRDPASSENPELLMFPVSYPGLRQRKDWIIPSCPENKEGFPNQLVQKSNRDKET 188
QY 140 KIFYSITGPGADSPGEPGFAVEKETGWLKLLKPLDREIAKYELFGHVSNGASVEDPM 199
DB 189 KVFYSITGQADKPPVGVFTIERETGWLKVTQPLDREIAKYILYSHAVSSNGAEVDEPM 248
QY 200 NISITVDQNDHKPKFTQDFRGSVLEGLVPGTSMQVATDDEDDAIYINGVAYSIHS 259
DB 249 EIVITVDQNDNRPEFTQEVFEGSVAEGAVPGTSMQVATDDEDDAIYINGVAYSIHS 308
QY 260 QEPKDPHDLMTIHRSTGTTISVSSGLDREKVEPYITITQATDMDGSGSTTTAVAVVEIL 319
DB 309 QDPKPEKMTVNRDGTGLSVLTSLGDRSYFTYTLVQADLQSGSLSTAKAVITVK 368
QY 320 DANDNAPMPQKYEAHVHPNACVHEVQRLITVTDLPAPNSAPKARYILMGDGDGHPIT 379
DB 369 DINDNAPVFNSTYQGVPEVNEVNIATLKVTDDEDDAPNTEAKVYTVV-NDPDQGVV 427

RESULT 6

S34438

uvomorulin - mouse

C:Species: Mus musculus (house mouse)

C:Date: 22-Nov-1993 #sequence_revision 03-Nov-1995 #text_change 23-May-1997

C:Accession: S34438

R:Ringwald, M.; Baribault, H.; Schmidt, C.; Kemler, R.

Nucleic Acids Res. 19, 6533-6539, 1991

A:Title: The structure of the gene coding for the mouse cell adhesion molecule uvomorulin

A:Reference number: S34438; MUID:92093614; PMID:1754391

A:Accession: S34438

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-884 <RIN>

A:Cross-references: EMBL:X60975

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1992

C:Genetics:

A:Introns: 16/3; 57/1; 131/3; 179/3; 231/3; 280/1; 338/3; 381/3; 442/3; 524/2; 573/1; 641/1

C:Superfamily: cadherin; cadherin repeat homology

F:380-488/Domain: cadherin repeat homology <CR3>

Query Match

Best Local Similarity 54.5%; Score 2379; DB 2; Length 884;

Matches 471; Conservative 122; Mismatches 222; Indels 66; Gaps 13;

QY 9 ASLLQLQV-CWLOCAASEP--CRAVFEAEVTLAAGAEQEPGALGV-FMGCPGQEP 64

DB 10 ALLLLQLQVSSWL-COELEPSCSPGFSSEVYTFPVEHLERGHVLRVFECTGRPT 68

QY 65 LFSTNDNDFTVRNGETVOERSLK----- 88

DB 69 AFPSDSRFKVTADGTTITVKRHLKHLKLETSFLVRARDSSHRELSTKVTLSKMGHHHRH 128

QY 89 -----ERNP-LKIPSKRI-LRRHKRDWVAPISVPENKGGPPQRLNOLSKNDRDT 139

DB 129 HHRDPASSENPELLMFPVSYPGLRQRKDWIIPSCPENKEGFPNQLVQKSNRDKET 188

QY 140 KIFYSITGPGADSPGEPGFAVEKETGWLKLLKPLDREIAKYELFGHVSNGASVEDPM 199

DB 189 KVFYSITGQADKPPVGVFTIERETGWLKVTQPLDREIAKYILYSHAVSSNGAEVDEPM 248

QY 200 NISITVDQNDHKPKFTQDFRGSVLEGLVPGTSMQVATDDEDDAIYINGVAYSIHS 259

DB 249 EIVITVDQNDNRPEFTQEVFEGSVAEGAVPGTSMQVATDDEDDAIYINGVAYSIHS 308

QY 260 QEPKDPHDLMTIHRSTGTTISVSSGLDREKVEPYITITQATDMDGSGSTTTAVAVVEIL 319

DB 309 QDPKPEKMTVNRDGTGLSVLTSLGDRSYFTYTLVQADLQSGSLSTAKAVITVK 368

QY 320 DANDNAPMPQKYEAHVHPNACVHEVQRLITVTDLPAPNSAPKARYILMGDGDGHPIT 379

DB 369 DINDNAPVFNSTYQGVPEVNEVNIATLKVTDDEDDAPNTEAKVYTVV-NDPDQGVV 427

```

Db 189 KVFYSITGQADKPPVGVFIIERETGWLKVTPQLDREAIAKYILYSHAVSSNGEAVEDEPM 248
QY 200 NISIIVTQNDHKKPKFTQDFRGSVLEGLVPGTSVMQVATDEDDALYYNGVWVAYSIIHS 259
Db 249 EIVITVTQNDNRPEFTQPFEGFAEGVGTSMVKVSATDADDVNTYNAIAIATVVS 308
QY 260 QEPKDPDLFTTHRSCTTSSVSSGLDRKVPYETLTIQATMDGDDGSGTTTAVAVEIL 319
Db 309 QDPFLPHKNMFTVNRDFGVISLTSGLDRSPYPTIYLWQAADLQEGSLTAKAVITVK 368
QY 320 DANDNAPMFPQKYEAVHPENAVGHEVQRLVTVTDLPNSPAWRATYILMGDDGDPHFI 379
Db 369 DINDNAPFNPSTYQGVQVENEVARIATLKTDADDAPNTPAWKAVTVV-DNPDQCFV 427
QY 380 TTHPESNOGILTTKGLDFAKNGHTLYVEVTNEAPVFLKPTSTATIVVHVEDVNEAPV 439
Db 428 VTDPTTNDGILKTAAGLDFAKQYILHVRVNEEPEGSLVSTATVTVDVVNEAPI 487
QY 440 FVPESKVEVQEGIPTGEPVCVYTAEDPK-ENQKISYRILRDPAGWLANDPDSGQVAV 498
Db 488 FMPAERRVEVPEDRGVQGEITSYTABRPTDFMDOKITRYIWRDTANWLEINPETGIFTR 547
QY 499 GTLDREDEQFVNNIYVWVLANMGSPPTGTGTLTLLIDVNDHGPVPEPQIITCNQ 558
Db 548 AEMDREDAEHVKNSTYVALIATDGGPIATGTGTLTLLVLDVNDNAPIPEPRNQFCQR 607
QY 559 SPVRHLNITDKLSPTSPPQAOLTDSDIYNTAEVNEEG-DTVVLSLKKFLKQDPTDV 617
Db 608 NPQPHIITLDPDLPNTSPETAELTHGASWNITIEYNDAAQESLLQPRKDLIEGYKI 667
QY 618 HLSUSDHNGKQLTVIRATVCDCHGVETCPGPKGKF-----ILPVLGAVLALLF 668
Db 668 HLKLDNQNKQVTLTDVHVDCSGTVNVC---MKAGIVAAGLQVPAILGILGILALLI 724
QY 669 LLLVLLLVLRKKIKPRLPLPEDTRDNVYFYEGEGGEEDQYDITQLHGLEARPEV 728
Db 725 LILLLLFLRRTVVKFLPFPDDTRDNVYFYDEEGGEEDQDFLSQLHGLDARPEV 784
QY 729 VLRNDVAPTIPTMYRPRANPEIGNFIENUKAANTPTAPPYDTLLVDFYEGSGSD 788
Db 785 T-RNDVAPTLMSVQYRPRANPEIGNFIDENUKAADPTAPPYDLSLLVDFYEGSGSE 843
QY 789 AASLSLTSASDQDDYDYLNEGSRFKKLADMYGGEDD 829
Db 844 AASLSLTSASDQDDYDYLNEGSRFKKLADMYGGEDD 884

RESULT 7
LJXLCP
EP-cadherin precursor - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C>Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C:Accession: B43785; A60128
R:Ginsberg, D.; DeSimone, D.; Geiger, B.
Development 111, 315-325, 1991
A:Title: Expression of a novel cadherin (EP-cadherin) in unfertilized eggs and early xen
A:Reference number: A43785; MUID:91372132; PMID:1893866
A:Accession: B43785
A:Molecule type: mRNA
A:Residues: 1-895 <GIN>
A:Cross-references: GB:X63720; NID:g64681; PIDN:CAA45252.1; PID:g64682
A:Note: it is uncertain whether Met-1 or Met-16 is the initiator
R:Angres, B.; Mueller, A.H.J.; Kellermann, J.; Hausen, P.
Development 111, 829-844, 1991
A:Title: Differential expression of two cadherins in Xenopus laevis.
A:Reference number: A60128; MUID:91347911; PMID:1879345
A:Accession: A60128
A:Molecule type: protein
A:Residues: 171-177, 'I', 179-183, 'K', 185-189, 'X', 'ANG'
A:Note: the material sequenced may have contained U-cadherin as well as E-cadherin
C:Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought
C:Superfamily: cadherin; cadherin repeat homology
C:Keywords: calcium binding; cell adhesion; duplication; embryo; glycoprotein; transmembr

```

```

F:1-43/Domain: signal sequence #status predicted <SIG>
F:44-170/Domain: propeptide #status predicted <PRO>
F:171-895/Product: EP-cadherin #status predicted <MAT>
F:171-718/Domain: extracellular #status predicted <EXT>
F:173-278/Domain: cadherin repeat homology <CR1>
F:248-253/Region: cadherin binding #status predicted
F:281-391/Domain: cadherin repeat homology <CR2>
F:394-502/Domain: cadherin repeat homology <CR3>
F:503-610/Domain: cadherin repeat homology <CR4>
F:611-714/Domain: cadherin repeat homology <CR5>
F:719-743/Domain: transmembrane #status predicted <TM>
F:744-895/Domain: intracellular #status predicted <INT>
F:852-865/Region: serine-rich
F:440.696/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match          54.4%; Score 2375; DB 1; Length 895;
Best Local Similarity 54.2%; Pred. No. 1.7e-143;
Matches 471; Conservative 128; Mismatches 214; Indels 56; Gaps 9;

QY 11 LLLQVCMWQCAASEPCRAVFRBAEVILEAGCAEQEQCALGKV-EMGCCPQEQEPALFSTD 69
Db 32 LCLLQVPSINADVSGCKPGFSAEYIFSVNRRELERGRKLGKVFNSDCTTRKGLDVG 91
QY 70 NDDFTV-----NNGETVQERRS 86
Db 92 DSRFVLPDGTIVLKVHVKLHKDKFTISTWDARGIKHSTNIAVASKRHRSGEASRGS 151
QY 87 LKERNPLKIFPSKRI-LRRHRRDWVAPIVSPENKGGKPPFORLNQLKSKNDRTKIYSI 145
Db 152 ---KLPLVITPETHETGLKRRKRDWIPPIKVSNERGEPFKRLVQIKSKDRFNKYYSI 208
QY 146 TGGGADSPGGEVFAVEKETGMLLNKPLDREBIKAKYELFGHAVSENGASVEDPMNISIIV 205
Db 209 TQGGADNPQGVFRLEWETGHWLTVPLDRREYDKYVLSHAVSENGSPVEEPMEIINV 268
QY 206 TDQNDHKPKFTQDTPRGSVLEGLVPGTSVMQVATDEDDALYYNGVWVAYSIIHQEPKDP 265
Db 269 IDQNDNRPKFTQDTPRGSVLEGLVPGTSVMQVATDEDDALYYNGVWVAYSIIHQEPKDP 328
QY 266 HDLMEIHRSTGTSIVSSGLDRKVPYETLTIQATMDGDDGSGTTTAVAVEILDANDNA 325
Db 329 INPLFTINRETVISLIGTGLDRKVPYETLTIQATMDGDDGSGTTTAVAVEILDANDNA 388
QY 326 PMFDPQKYEAVHPENAVGHEVQRLVTVTDLPNSPAWRATYILMGDDGDPHFIITTHPES 385
Db 389 PIFDEKTYTALVPENEIGFEVQRLSVTDLDMPGTPAWQAVYKIR-VNEGGGFNITDPES 447
QY 386 NOGILTKRGLDFAKNGHTLYVEVTNEAPVFLKPTSTATIVVHVEDVNEAPVPEPSK 445
Db 448 NOGILTKRGLDFAKNGHTLYVEVTNEAPVFLKPTSTATIVVHVEDVNEAPVPEPSK 507
QY 446 VVEVEGPIPTGEPVCVYTAEDPKEN-QKISYRILRDPAGWLANDPDSGQVAVTGLDRE 504
Db 508 RVDSSEDLRSGEKIISLVAQDPDKQIQKLSYFIGNDPFWLTVNKGNGIVTGNGLDRE 567
QY 505 DEQFVNNIYVWVLANMGSPPTGTGTLTLLIDVNDHGPVPEPQIITCNQSPVRYH 564
Db 568 SE-YVKNNTYTVIMLVTDGVSQVGTGTLILHVLVDNDNGVPPEPRVFTWCDQNPQV 626
QY 565 LNTDKLSPTSPPQAOLTDSDIYNTAEVNEEGDVTVLKPKKQDQYDVLHLSLDH 624
Db 627 LTISADIPNTPYKVSLSHSGSDITWAKELDSKTSMLLSPQQLKKGDSYIYVLLSDA 686
QY 625 GNKEQLTVIRATVCDCHGVETCPGFWKGGFTLP-----VLGAVLALLFLLVLLVLRKK 680
Db 687 QNPNQTVVNVATVCSGEGKAIKQEKLVGGFDLPILVILGSLVALLILFLLVLLFLKRX 746
QY 681 RKIKEPILLPEDTRDNVYFYEGEGGEEDQYDITQLHGLEARPEVVLVNDVAPTIIP 740
Db 747 KVKKEPLLPEDDTRDNVYFYEGEGGEEDQYDYSQJHGLDSPD-IMENDVVPILMP 805
QY 741 TPMYRPRANPEIGNFIENUKAANTPTAPPYDTLLVDFYEGSGSAAASLSLTSAS 800

```


Db 806 ABHYPRPNPDEIGNFIDENLDAANDPTAPPYDLSLLVFDYEGSGEAAASLSSINSNS 865
 QY 801 DQDQDYDLNMGWSRFXKLADWYGGGDD 829
 Db 866 NDEHDYNYLSDWGSFRKLADWYGGDDDE 894
 RESULT 8
 S43064
 cadherin - African clawed frog
 C:Species: Xenopus laevis (African clawed frog)
 C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 20-Aug-1999
 C:Accession: S43064; S43065
 R:Kuehl, M.
 submitted to the EMBL Data Library, March 1994
 A:Reference number: S43064
 A:Accession: S43064
 A:Molecule type: mRNA
 A:Residues: 1-905 <KUB>
 A:Cross-references: EMBL:X78546; NID:q468816; PIDN:CAA55292.1; PID:g468817
 R:Herzberg, F.; Wildermuth, V.; Wedlich, D.
 Mech. Dev. 35, 33-42, 1991
 A:Title: Expression of Xscad, a novel cadherin, during oogenesis and early development
 A:Reference number: S43065; MUID:92062581; PMID:1840622
 A:Accession: S43065
 A:Molecule type: mRNA
 A:Residues: 'NSA', 462-697, 'O', 699-807, 'A', 809-840, 'N', 842-877, 'N', 879-883, 'N', 885-902, 'E'
 A:Cross-references: EMBL:X78546
 C:Superfamily: cadherin; cadherin repeat homology
 C:Keywords: calcium binding; cell adhesion; glycoprotein; transmembrane protein
 F:181-286/Domain: cadherin repeat homology <R1>
 F:289-399/Domain: cadherin repeat homology <R2>
 F:402-510/Domain: cadherin repeat homology <R3>
 F:511-618/Domain: cadherin repeat homology <R4>
 F:619-722/Domain: cadherin repeat homology <R5>
 Query Match 54.3%; Score 2371; DB 2; Length 905;
 Best Local Similarity 54.1%; Pred. No. 3,1e-143;
 Matches 469; Conservative 125; Mismatches 222; Indels 51; Gaps 8;
 QY 11 LLLQVWCWLQCAASEPCRAVFREAETVLEAGGAQEFGQALGV-FMGCPGQEPALFSTD 69
 Db 39 LCLLQVPSINVDVSGGQPGFSSNYYFVSNNRELRGRKLGNLVDCTTRKHGUYDVG 98
 QY 70 NDDFTVRNGTV-----QERRSLK 88
 Db 99 DSRFVLPDGTNLVKRWKHLKSKDTRFTISTDARGIKHSTNISVNVKGRHSGEARSRS 158
 QY 89 ERNPLKTFPSKRI-LRRHKDWWAPISVPEGKGPFPQRLNQLKSKORDYKIFYSITG 147
 Db 159 SELPVLFPPEKHTGLKGRKRDWIPPIKVSNERGEPFPKRLVQIKSKNKLKSVFFSITG 218
 QY 148 PGADSPPEGVFAVEKETGWLKLLNKLPLDREIAKYLEFGHVAVSNGASVEDPMNISIIVTD 207
 Db 219 QGADTPPEGIFRIEKEGTGMQVTRPLDREYEVKLVLLSHAVSNGASVEEPEMITVTVID 278
 QY 208 QNDHKPKFTQDTRGSLVGLVPGTSMQVATDEDDAIYTYNGVAVYAIHQEPKDPHD 267
 Db 279 QNDNRPAFTQPVFRGSRVREGVQGTGKMSVSAITDDSDSLNGVIAYSILKQDPEEPT 338
 QY 268 LMFTIHRSTGTISVISGLDREKVPYETLTIQATDMDGSGTITTAVAVVEILLDANDNAPM 327
 Db 339 NLFTINRETGVISLIGTGLDREKFPETLTQADLDGAGLTAEGRKAVIEITDANDNAPI 398
 QY 328 FDPQKYEAPVPEAVGHEVQRLLVTDLDPNSPAPWATYILMGDDGDGHTFTTHPESNQ 387
 Db 399 FDPKTYTALVPEVEVGEVQRLLSVTDLMDPGTAAQVAKIR-VNEGGEFNNITDTPESNQ 457
 QY 388 GILTTTRKGLDFAEKNOHTLVVEVNEAPFVLKLPSTATIVVHVEDVNEAPVFPSPKV 447
 Db 458 GILTTAKGLDFEVRKQVIOITVENAVPSVPLPTSTATVTVTVEDVNEAPVFPVRSV 517
 QY 448 EVEGEPITGPVPCVYTAEDDPDEN-QKISVRIILRDPAGWLAMPDPSGQVATVGLDREDE 506

Db 181 LATDGSPTTGTGTLTLLTMDINDHGPVPEPQITCNQSPVQVNLNITDKLSHTAP 240
 Qy 579 FOAQLTDDSDIYTAENEGEDTVVLSLKKFLKQDYDVHLSLSDHGNKEQLTVIRATVC 638
 Db 241 FOAQLTHSDSVYTAENEGEDVALSLKKFLKQGEYDVHLSLSDHGNKEQLTVIRATVC 300
 Qy 639 DCHGVETCPGPKGGFLLVGLAVLALLFLLVLLVLLVLRKKKIKPELLEDDTRDNV 698
 Db 301 DCHGNVTCRDPWTGFLFLLGALLLALLVLLVLRKKKIKPELLEDDTRDNV 360
 Qy 699 FYGEGGGEDQDYDTQLHRLGLEARPEVLRNDVAPTIPTMYRPRPANDDEIGNFI 758
 Db 361 FYGEGGGEDQDYDTQLHRLGLEARPEVLRNDVAPSIPTMYRPRPANDDEIGNFI 420
 Qy 759 IENLKAANTDPTAPPYDTLAVFYEGSGDAASISSLTSSASQDQDYDYLNEWGSRPKK 818
 Db 421 IENLKAANTDPTAPPYDLSLAVFYEGSGDAASLSSLTSSDQDQDYDYLNEWGSRPKK 480
 Qy 819 LADMYGGGDD 829
 Db 481 LADMYGGGDD 491

RESULT 10
 IUCHCL
 B-cadherin precursor, hepatic - chicken
 N:Alternate names: L-CAM; liver cell adhesion molecule
 C:Species: Gallus gallus (chicken)
 C:Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 25-Oct-1996
 C:Accession: A30201; A29866; B29866
 R:Sorkin, B.C.; Hemperly, J.J.; Edelman, G.M.; Cunningham, B.A.
 Proc. Natl. Acad. Sci. U.S.A. 85, 7617-7621, 1988
 A:Title: Structure of the gene for the liver cell adhesion molecule, L-CAM.
 A:Reference number: A30201; MUID:89017248; PMID:3174655
 A:Accession: A30201
 A:Molecule type: mRNA
 A:Residues: 1-81 <SOR>
 A:Cross-references: EMBL:J04074
 R:Gallin, W.J.; Sorkin, B.C.; Edelman, G.M.; Cunningham, B.A.
 Proc. Natl. Acad. Sci. U.S.A. 84, 2808-2812, 1987
 A:Title: Sequence analysis of a cDNA clone encoding the liver cell adhesion molecule, L-
 A:Reference number: A29866; MUID:87204217; PMID:3472238
 A:Accession: A29866
 A:Molecule type: mRNA
 A:Residues: 51-887 <GAI>
 A:Cross-references: EMBL:M16260
 R:Accession: B29866
 A:Molecule type: protein
 A:Residues: 161-172;323-336;386-407;533-551 <GA2>
 C:Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought
 C:Genetics:
 A:Introns: 23/3; 62/1; 138/1 183/3; 235/3; 284/1; 342/3; 385/3; 447/3; 529/2; 577/1; 652/1
 C:Superfamily: cadherin; cadherin repeat homology
 C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; liver; transmembr
 F:1-26/Domain: signal sequence #status predicted <SIG>
 F:127-160/Domain: propeptide #status predicted <PRO>
 F:161-887/Product: E-cadherin, hepatic #status experimental <MAT>
 F:161-704/Domain: extracellular #status predicted <EXT>
 F:163-268/Domain: cadherin repeat homology <CR1>
 F:163-243/Region: cadherin binding #status predicted
 F:271-381/Domain: cadherin repeat homology <CR2>
 F:384-493/Domain: cadherin repeat homology <CR3>
 F:494-601/Domain: cadherin repeat homology <CR4>
 F:602-704/Domain: cadherin repeat homology <CR5>
 F:705-735/Domain: transmembrane #status predicted <TM>
 F:736-887/Domain: intracellular #status predicted <INT>
 F:844-857/Region: serine-rich
 F:291,346,564,643/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 52.0%; Score 2274; DB 1; Length 887;
 -Best Local Similarity 52.7%; Pred. No. 4,6e-137;
 Matches 461; Conservative 122; Mismatches 236; Indels 55; Gaps 13;

Qy 8 LASLILLQVCMWLOCAASEPCRAVFEAEVLEAGABEQEPGQALGV-FMCGPQGPALF 66
 Db 16 LVLLILLQVCGRRCDAAAPQPGFAETFSFSPQDSVAAGRELGRVFAACSGRPWAV 75
 Qy 67 -STD-----NDD-----FTVANGETVQERRS----- 86
 Db 76 VPTDRFRKVGNDGVVSKRPLTLYGRKISITIIYAQDANGKHSARVTVGRHRRHHNH 135
 Qy 87 -LKERNPLKI-FP--SKRIIRRRHRDWWVAPISVPENKGFPPQRLNQLKSNKRDRTKIF 142
 Db 136 HLQDTTAVLTFFKHPDGLRRQRQKRDVPIPSIENHERGPPMRLVQIKSNKDKESKV 195
 Qy 143 YSITGPGADSPGEGVFAVEKETGMLLNKLPDLDRREIAKYLFGHAVSENGASVEDPNIS 202
 Db 196 YSITGQGADSPVGIFITIERETGWLVEVTEQDLREKIDRYTLLSHAVASQGPVEDPMEI 255
 Qy 203 IIVTDQNDHKFKTQDTFRGSVLEGVLPGTSVMQVTTATDEDDAIYTVNGVVAYSIHQEP 262
 Db 256 ITVMDQNDKVPFIKEVFGVIEENAKPGISVMIVNATDADDAVNTDNGIVSYIVSQP 315
 Qy 263 KDPHDLMTIHRSTGTSIVSSGLDRKVPETITIOATMDGSGSTTAVAVVEILDAN 322
 Db 316 PRPHQMETIDPAKGIISVLGTLDRETTNYTLIVQATDOEGKGLSNTAIIIEVDAN 375
 Qy 323 DNAPMFDPOKYEAVPENAVGHEVQRLTVTDLDPNSPAMRATYILMGDGDGDRFTITTH 382
 Db 376 DNIPIFNMTGEGVVEENKPGTEVARLTVTDQDAPGSPAWQAVYHKSGLNDGAFSIITD 435
 Qy 383 PESNGILTTTRKGLDFAKNQHTLYVEVNEAPVFLKLPSTATIVVHVVDNEAPVFP 442
 Db 436 PSTNGILLKTAAGLDYETKSYDLVVTVENKPLSVPIITLSTASVLTVLDWNEPFPVP 495
 Qy 443 PSKVVEGEGPTGEPVVCVYTAEPDKE-NQKISVRIILDPAGLAMPDSDGQVTAAGTL 501
 Db 496 PIKRVGVPEDLPVGQCVTSYTAQDPDRMRQKITIRMGSDDPAGWLYHPENGIIVTAQPL 555
 Qy 502 DREDEQFVANNIYEVWVLAMNDGSPPTTGTGLTLTIDVNDHGVPEPPEPPIITCNOSPV 561
 Db 556 DRESVHAI-NSTYKAIILAVDNGIPDTTGTGLTLLQDNDVNDGPTPEPRSFCSRQPE 614
 Qy 562 RHVLNITDKLSPTSPQAQLTDDSDIYTAENEGEDTVVLSLKKFLKQDYDVHLSL 621
 Db 615 KQILSVLDKLPSPHTYFPFKAALHSGSSNNWTVIRGQ-DELAGMLKKELECEGINFVKL 673
 Qy 622 SDHGNKQSLTVIRATVCCHGHVETCP--GPWKGGF----ILPVLGAVIALLLLLVLL 675
 Db 674 TDSQKAQVTVKAQVCEGCTAKNCERRSYVGLGVPAILGILGILALLILLILL 733
 Qy 676 LVKRRKIKPELLEDDTRDNVFFYEGGEGEDQDYDTQLHRLGLEARPEVLRNDVA 735
 Db 734 FARRKKEKEPLLPEDDMRNVNVDYEGGEGEDQDYDLSQLHRLGLDARPEVI-RNDVA 792
 Qy 736 PIITPTMYRPRPANDDEIGNFIENLKAANTDPTAPPYDLSLVDYEGSGSDAASLSSL 795
 Db 793 PPLMAQPRPAPNDDEIGNFIDENLKAANTDPTAPPYDLSLVDYEGSGSEATSLSSL 852
 Qy 796 TSSASDQDQDYDYLNEWGSRPKKLADMYGGGDD 829
 Db 853 NSSASDQDQDYDYLNEWGNRFPKKLAELVGGGDD 886

RESULT 11
 S47518
 cadherin - African clawed frog
 C:Species: Xenopus laevis (African clawed frog)
 C:Date: 01-Feb-1995 #sequence_revision 10-Nov-1995 #text_change 17-Mar-1999
 C:Accession: S47518
 R:Tooi, O.; Fujii, G.; Tashiro, K.; Shiohawa, K.
 Biochim. Biophys. Acta 1219, 121-128, 1994
 A:Title: Molecular cloning of cDNA for XTCD-1, a novel Xenopus cadherin, and its express
 A:Reference number: S47518; MUID:94368839; PMID:8086449
 A:Accession: S47518
 A:Molecule type: mRNA

A;Residues: 1-871 <TO>
C;Superfamily: cadherin; cadherin repeat homology
F;151-256/Domain: cadherin repeat homology <CR1>

Query Match
Best Local Similarity 50.2%; Score 2192.5; DB 2; Length 871;
Matches 460; Conservative 126; Mismatches 227; Indels 73; Gaps 16;

QY 1 MGLPR-----GFLASLLLOVC-----WLQCAASEPCRAVFRABVTLEAGABEQPQAL 51
DB 1 MGLRPMLLGAVLLITLQVGLAEWTQ-----CRNGFSKESYFLV-PKNLETKAL 53

QY 52 GKV-FMGCPCGEPALFSDNDFTVRNGETVOERSLKERNPLKF----- 96
DB 54 GRVFNCEGVRVQFASKOPNFHKGDTYIIKNPARKMDNRKTRFVLAWETKGVYST 113

QY 97 -----PSKILRRHRKRDWVAVPISVPENKGPFPFQRLNQLK 132
DB 114 NITLKREGRHRQDLFSGKSHHPKSETGLKQKQKRDWVIPPILVSENEKGPFPKRIQIK 173

QY 133 SNKDRDTKIFYSIITPGADSPPEGVFAVEKETHLLNKLPLDRBIAKYELFHAVSENG 192
DB 174 SSYAKEVKVYSITGGADTFPEGVFAIGREDGLWLVTRPLDREADNYLVLFHAVSSNG 233

QY 193 ASVEDPMNISIIITDQNDHKPKFTQDTFRGSLVLEGLVPGTSMQVATDDEDAIYTVNGV 252
DB 234 ANVEDPMELIKVQDQNDNDVFTQSVFEGSPESKPGTAVMTVSATDADSDVDYNGV 293

QY 253 VAYSIHQEPKDPHLMFTTHRSCTTSSVSSGLDRKVPZEYTLTIQATDMD-DGSGTTT 311
DB 294 ITYSILNQEPKEPTNKNFTTHSESGLISVLTTGLDREKNPVYTLTIQAADGEGKORTTT 353

QY 312 AVAVVEILDANDAMPDPKYEAAHVPENAVGHVQRLTWTDDAPNSPAWRATYLLMG 371
DB 354 ATALIVMDTNDNPVDFPTQYAKVENEVGHVAVLTVDIEDIGTDANNAVYKLIK 413

QY 372 DGGHFTTTHPSNQSLTRKGLDPEAKNQHTLYVEVNEAPFVLKLTSTATIVVHY 431
DB 414 NEANYFSIQT-DTGNIGLLTKVGLDYELKQYILSVITNKANFSVPLQTSTATVTVS 472

QY 432 EDVNEAPVFPVPPSKVVEQCIPTGEPVCVYTAEDPKK-NOKISYELILDPAWLAMP 490
DB 473 EDVNEAPFIFPPVKEVSSDLPSGQVAVTYTAQDPDKQNKITYVIGNDPGRVSVNK 532

QY 491 DSGQVAVGLDREDEQFVNNIYVVMVLAMNGSPPTTGTGTLILLTLIDVNDHGPVPEP 550
DB 533 DNGIVTGNGLDRE-SKFLVNNYKVILLAADSGSPSATGTGLVNLNLDVNDNGFLEP 591

QY 551 RQITICNQSPVRHVLNTDKDLSPTSPFOQLTDDSDIYVTAEVNEEGTVVLSL--KK 608
DB 592 QSESFCQKPCFRVFTIIDRLSPNTYFYKAELTGESNENWTAIVTDK-----ILELRPKK 647

QY 609 FLKQDTYDVHLSLSDHGNKQLTVIRATVCDCHGVETC--PGPWKGGF-----ILPVLGA 662
DB 648 ELBIGQVDMITLSDSLGNVTKLHTICQGDGRKQCEKAAIAGGLGISAVGLGG 707

QY 663 VLALLFLLVLLVLRKKRIKPLLPEDDTRDNVYFEGEGGEDQDYITQLHRLG 722
DB 708 ILALLLLLLLLLVRKKVVKPEPLLPEDDTRDNVFSYDEEGGEDQDFDLSQLHRLG 767

QY 723 EAPPEVLRNDVAPTIPTPWRPRPANDEIGNFIENLKANTDPTAPPDYTLVIFY 782
DB 768 DAPPDVT-RNDVAP-VLAAQYPRPANDEIGNFIENLNADNDPTAPPDYSLVIFY 825

QY 783 EGSQSDAASLSSTSSASDQDQDYVLENGSRFKKLADMYGGED 828
DB 826 EGSQSEAASSLSSNSNDDLQDYSALNDWNGPFTKLADMYGGED 871

RESULT 12
1JHUCN
cadherin 2 precursor - human
N;Alternate names: N-cadherin; neuronal cadherin

C;Species: Homo sapiens (man)
C;Date: 30-Jun-1993 #sequence revision 30-Jun-1993 #text_change 22-Jun-1999
C;Accession: A38870; S11487; JQ0751; S13799
R;Reid, R.A.
Submitted to the EMBL Data Library, November 1990
A;Reference number: A38870
A;Accession: A38870
A;Molecule type: mRNA
A;Residues: 1-906 <RE1>
A;Cross-references: EMBL:X54315; NID:G34998; PIDN:CAA38213.1; PID:G34999
R;Reid, R.A.; Hemperly, J.J. 1990
Nucleic Acids Res. 18, 5896, 1990
A;Title: Human N-cadherin: nucleotide and deduced amino acid sequence.
A;Reference number: S11487; MUID:91016946; PMID:2216790
A;Accession: S11487
A;Molecule type: mRNA
A;Residues: 1-340, 'N', 342-698, 'R', 700-704, 'F', 706-906 <RE2>
A;Cross-references: EMBL:X54315
A;Note: this sequence has been revised in reference A38870
R;Walsh, F.S.; Barton, C.H.; Putt, W.; Moore, S.E.; Kelsell, D.; Spurr, N.; Goodfellow, I.
J. Neurochem. 55, 805-812, 1990
A;Title: N-cadherin gene maps to human chromosome 18 and is not linked to the B-cadherin
A;Reference number: JQ0751; MUID:90347462; PMID:2384753
A;Accession: JQ0751
A;Molecule type: mRNA
A;Residues: 160-194, 'IR', 197-211, 'L', 213-227, 'Q', 229, 'N', 231-235, 'G', 237-248, 'T', 250-356,
A;Cross-references: GB:M34064
C;Comment: Cadherins mediate calcium-dependent intercellular adhesion and are thought to
C;Genetics:
A;Gene: GDB:CDH2; NCAD
A;Cross-references: GDB:128185; OMIM:114020
A;Map position: 18q12.1-18q12.1
A;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane prot
F;1-27/Domain: signal sequence #status predicted <SIG>
F;28-159/Domain: propeptide #status predicted <PRO>
F;160-906/Product: N-cadherin #status predicted <N-cadherin>
F;160-714/Domain: extracellular #status predicted <EXT>
F;162-267/Domain: cadherin repeat homology <CR1>
F;237-242/Region: cadherin binding #status predicted
F;270-382/Domain: cadherin repeat homology <CR2>
F;385-497/Domain: cadherin repeat homology <CR3>
F;500-605/Domain: cadherin repeat homology <CR4>
F;606-712/Domain: cadherin repeat homology <CR5>
F;715-746/Domain: transmembrane #status predicted <TM>
F;747-906/Domain: intracellular #status predicted <INT>
F;865-878/Region: serine-rich
F;190,273,325,402,572,622,651,692/Binding site: carbohydrate (Asn) (covalent) #status pre

Query Match 39.1%; Score 1707; DB 1; Length 906;
Best Local Similarity 42.7%; Pred. No. 7,3e-101;
Matches 390; Conservative 129; Mismatches 292; Indels 102; Gaps 27;

QY 6 GPLASILLLOVCWLOCAASEP-----CRAVFRABVTLEAGABEQPQALGKV-FMGCP 59
DB 6 GALTLLPLLALLOALQVSAESGAICTKGFPE-DVYSAVLKDVHEGQPLINXVFN 64

QY 60 GQ-----BPALFSTNDND--FTVRN-----GETVOER----- 84
DB 65 GKRKQVYESSBPADFKVDEGDMVAVRSFPLSSSEHAKFLIYAQDKETQKQVAVKLSLK 124

QY 85 -----RSJKERNPLK--IFP---SKRI---LRHRKRDWVAVPISVPENKGPFPQRLNQLK 132
DB 125 PTLTEESVKEAEVEEIVFPQFSKSHQKQKRDWVIPPILVSENEKGPFPQELVIR 184

QY 133 SNKDRDTKIFYSIITPGADSPPEGVFAVEKETHLLNKLPLDRBIAKYELFHAVSENG 192
DB 185 SDRDKSLRYTVTGPAGDQPTGTGIINPISGQSLVTKPLDREQIAHFHRAHVAIDNG 244

QY 193 ASVEDPMNISIIITDQNDHKPKFTQDTFRGSLVLEGLVPGTSMQVATDDEDAIYTVNGV 252
DB 245 NQVENPIDIVINVIDMNDNRPEFLHQVWNGVTPGSKGTYYMTVTAIDADDP-NALNGM 303

253 VAYSISQEPKDPHDLMTTHRSSTGTSISVSGLDREKVPYTLTQATDMOG---DGSST 309
 304 LRVIVSQAPSTPSNNFTINNETGDIITVAAGLDREKVOQYTLTQATDMEGIPYGLS 363
 310 TTAVALVEILLDANDANAPKDPQYEAHVHPENAVNGHEVORLTVTDLDPNSPAWATYLM 369
 364 NTATAVITVDVNDNPPFTAMTPYGVENRVDIIIVANLTVDKQPHTPANNAVYRLS 423
 370 GGGDGDHFTITTHPESQGLITTRKGLDFAKQHTLYVETNEAPVLKL---PTSTAT 426
 424 GGDPTGFAIOTDPSNGLVTVVKPIDFETNRMFVLTAAENQVPLAKGIQHPQSTAT 483
 427 IVGVHVEDNEAPVFPVPSKVEVOEGIPTGEPVCVYTAEDPK-ENOKISYRILRDPAGW 485
 484 VSVTVIDVNEPFPAPFKIIRQEGELHAGTMLTFTTAQDPDRYMQNIRYTKLSDPANW 543
 486 LAMPDPSQVTAAGTLDREDEBQFVRNNIYEVVLMADNGSPPTTGTGTLTLLTLDVNDHG 545
 544 LKIDPVNGQITTIIVLDRESN-VKNNIYNATFLASDNGIPWMSGTGTLOIYLLDINDNA 602
 546 PVPSPRQITICNOSPVRHVNIT--DKDLSPTSPFOAQLTDDSDIY---WTAEVNEEGD 600
 603 PQVLPOEAETC-ETPDPNSINITALDYIDFNAGPFAFDLPFGVTIKRNTI-TRLNGD 660
 601 TVVLISLK-KFLKQDTYVHLSLSDHGN--KEQLTVIRATVCDCHG-----VETCPGPKW 652
 661 FAQLNLKIKFLEAGIYEVPIITDSGNPKSNISILAVKVCQDSNGCDCTDVRIVGAGL 720
 653 GGFILPVLGAVLALLFLLVLLLV-----RKRKIKPELPLLPEDDCTDNDVFIY 702
 721 G-----TGAIITALLCIILVLMFVVMKRDKERQAKQLLIDPEDDVRDNLKYD 774
 703 BEGGGEEDDYDITQLHGLEAREV-----VLRNDVAPTIIPMYRPRPA--NPDEIG 755
 775 BEGGGEEDDYDLQLOQPTVEPDAIKPVGIRNDRP-IAHAPQYVRSAAHPFGDIG 833
 756 NFITENKAANTDPTAPPYDITLLVFDYEGSGSDAASLSLTSASDQDQDYDLNENGSR 815
 834 DFINEGLKAANDPTAPPYDITLLVFDYEGSGTAGSLSSLSNSSSGEGEQDYDLNDWGR 893
 816 FKKLADMYGGGED 828
 894 FKKLADMYGGGD 906

RESULT 13
 IUCHGN
 N-cadherin precursor, neuronal - chicken
 C:Species: Gallus gallus (chicken)
 C:Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
 C:Accession: A29964
 R:Hatta, K.; Nose, A.; Nagafuchi, A.; Takeichi, M.
 J. Cell Biol. 106, 873-881, 1988
 A:Title: Cloning and expression of cDNA encoding a neural calcium-dependent cell adhesion molecule
 A:Reference number: A29964; PMID:86153917; PMID:2831236
 A:Accession: A29964
 A:Molecule type: mRNA
 A:Residues: 1-912 <HAT>
 A:Cross-references: GB:X07277; NID:G63649; PIDN:CAA30258.1; PID:G63650
 C:Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought to be involved in a variety of cellular processes including cell adhesion, cell migration, and cell differentiation.
 C:Superfamily: cadherin; cadherin repeat homology
 C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane protein
 F:1-28/Domain: signal sequence #status predicted <SIG>
 F:29-164/Domain: propeptide #status predicted <PRO>
 F:165-912/Product: N-cadherin #status predicted <MAT>
 F:165-720/Domain: extracellular #status predicted <EXT>
 F:167-272/Domain: cadherin repeat homology <CR1>
 F:242-247/Region: cadherin binding #status predicted
 F:275-387/Domain: cadherin repeat homology <CR2>
 F:390-502/Domain: cadherin repeat homology <CR3>
 F:505-611/Domain: cadherin repeat homology <CR4>
 F:612-720/Domain: cadherin repeat homology <CR5>
 F:721-752/Domain: transmembrane #status predicted <TM>

F:753-912/Domain: intracellular #status predicted <INT>
 F:869-884/Region: serine-rich
 F:278,330,407,578,628,657/Binding site: carbohydrate (Asn) (covalent) #status predicted
 Query Match 39.0%; Score 1704.5; DB 1; Length 912;
 Best Local Similarity 42.6%; Pred. No. 1.1e-100;
 Matches 390; Conservative 132; Mismatches 284; Indels 109; Gaps 27;

QY 7 PLASLL--LQVWLOCASEP-CRAVPR-----AEVLEAGAGAEQPGQALGV-FWGC 58
 DB 14 PLALMLAALQAPKATCEDMLCKMGFPEDVHSAVRSVHG-----GQPLLNRVQSC 68
 QY 59 PQCEPALF-STDNDDFTV-----ENGETVQERR----- 85
 DB 69 DENRKIYGSSEPEDFRVGEDGVVVAERSFOLSAEPTEFVVSAARDKETQEBWQVKLTP 128
 QY 86 -----SLKERNPLK--IPFSKRI-----LRRHKRMVYAPISVPEGKGPQRLNOL 131
 DB 129 EPAFTGASEKQDKKIEDIIFPAQYKQSSHLKROKRDWIPPIINLPNSRGFFPFQELVRI 188
 QY 132 KSKNDKDTKIFYSITPGADSPPEGVFAVEKETGMLINKLPDREELAKYELFGHAYSEN 191
 DB 189 RSDRKSLSLAYSVTGPADQPTGIFILNIPISGLSVTKELDREQLASPLRAHADV 248
 QY 192 GASVEDPNISIIIVTDQNDHKPKFTQDTFRGSLVGLPGTSVMQVATDEDDAIYNG 251
 DB 249 GNQVENPIDVINVDMDNRPEFLHQVWNGTVPEGSKPGYVMTVTAIDADDP-NAQNG 307
 QY 252 VVAYSIHQEPKDPHDLMTTHRSSTGTSISVSGLDREKVPYTLTQATDMOG---GS 308
 DB 308 MLRYILQAPSPSPNNFTINNETGDIITVAAGLDREKVOQYTLTQATDMEGNPTVGL 367
 QY 309 TTTAVAVVEILLDANDANAPMFPQYEAHVHPENAVNGHEVORLTVTDLDPNSPAWATYLI 368
 DB 368 SNTATAVITVDVNDNPPFTAMTPYGVENRVDIIIVANLTVDKQPHTPANNAVY 427
 QY 369 MCGDGDHFTITTHPESQGLITTRKGLDFAKQHTLYVETNEAPVLKL---PTSTA 425
 DB 428 TCGDPTGQFTITLTDNSNDGLVTVVKPIDFETNRMFVLTAAENQVPLAKGIQHPQSTA 487
 QY 426 TTVHVEDNEAPVFPVPSKVEVOEGIPTGEPVCVYTAEDPKENQKIS--YKILRDP 483
 DB 488 TVSIIVIDVNESPFVFPNPKLVQEEGLLAGSMLTFTTARDPDRYMQTSLRYSKLS 547
 QY 484 GMLAMPDPSQVTAAGTLDREDEBQFVRNNIYEVVLMADNGSPPTTGTGTLTLLD 543
 DB 548 NMLKIDPVNGQITTTAVLDRE-SIYVQNNMYNATFLASDNGIPWMSGTGTLOIYLL 606
 QY 544 HGPVPEPRQITICNOSPVRHVNIT--DKDLSPTSPFOAQLTDDSDIY---WTAEVNEE 598
 DB 607 NAFQVNPKEATTC-ETLQPNANITAVDPDIDNPAGPFAFELPDSPPSIKRNWTI-V 664
 QY 599 GDTVVLISLK-KFLKQDTYVHLSLSDHGN--KEQLTVIRATVCDCH-----GHVET 650
 DB 665 GCHAQLSLRIFLEAGIYDVPITVITDSGNPHASSTSVLKVKVCQCDINGDCTDVR 724
 QY 651 WKGGFILPVLGAVLALLFLLVLLLV-----RKRKIKPELPLLPEDDCTDNDVFI 700
 DB 725 GUG-----TGAIITALLCIILVLMFVVMKRDKERQAKQLLIDPEDDVRDNLK 778
 QY 701 YGEEGGEEDDYDITQLHGLEAREV-----VLRNDVAPTIIPMYRPRPA--NPDE 753
 DB 779 YDEEGGEEDDYDLQLOQPTVEPDAIKPVGIRNDRP-IAHAPQYVRSAAHPFGD 837
 QY 754 IGFNFIENKAANTDPTAPPYDITLLVFDYEGSGSDAASLSLTSASDQDQDYDLNENG 813
 DB 838 IGFDFINEGLKAANDPTAPPYDITLLVFDYEGSGTAGSLSSLSNSSSGEGEQDYDL 897
 QY 814 SRFKKLADMYGGGED 828
 DB 898 PRFKKLADMYGGGD 912

RESULT 14
IQMSEN
N-cadherin precursor, neuronal - mouse
C:Species: Mus musculus (house mouse)
C>Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C/Accession: A32759; A46163
R: Miyatani, S.; Shimamura, K.; Hatta, M.; Nagafuchi, A.; Nose, A.; Matsunaga, M.; Hatta, S.; Miyatani, S.; Copeland, N.G.; Gilbert, D.J.; Jenkins, N.A.; Takeichi, M.
Proc. Natl. Acad. Sci. U.S.A. 89, 8443-8447, 1992
A>Title: Genomic structure and chromosomal mapping of the mouse N-cadherin gene.
A/Reference number: A32759; MUID:89346748; PMID:2762814
A/Accession: A32759
A/Molecule type: mRNA
A/Residues: 1-906 <MI>
A/Cross-references: GB:M31131; NID:g192327; PIDN:AAA37353.1; PID:g309125
R: Miyatani, S.; Copeland, N.G.; Gilbert, D.J.; Jenkins, N.A.; Takeichi, M.
Proc. Natl. Acad. Sci. U.S.A. 89, 8443-8447, 1992
A>Title: Genomic structure and chromosomal mapping of the mouse N-cadherin gene.
A/Reference number: A46163; MUID:92409532; PMID:1528849
A/Accession: A46163
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 839-906 <MI2>
A/Cross-references: GB:S45011; NID:g256010; PIDN:AAB23356.1; PID:g256011
A/Note: sequence extracted from NCBI backbone (NCBIN:113759, NCBI:P113760)
C/Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought to be involved in cell-cell adhesion.
C/Superfamily: cadherin; cadherin repeat homology
C/Keywords: calcium binding; cell adhesion; duplication
F:1-27/Domain: signal sequence #status predicted <SIG>
F:28-159/Domain: propeptide #status predicted <PRO>
F:160-906/Product: N-cadherin #status predicted <MAT>
F:160-714/Domain: extracellular #status predicted <EXT>
F:162-267/Domain: cadherin repeat homology <CR1>
F:237-242/Region: cadherin binding #status predicted
F:270-382/Domain: cadherin repeat homology <CR2>
F:385-497/Domain: cadherin repeat homology <CR3>
F:500-605/Domain: cadherin repeat homology <CR4>
F:606-714/Domain: cadherin repeat homology <CR5>
F:715-746/Domain: transmembrane #status predicted <TM>
F:747-906/Domain: intracellular #status predicted <INT>
F:865-878/Region: serine-rich
F:190,273,325,402,572,651,692/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 38.9%; Score 1697.5; DB 1; Length 906;
Best Local Similarity 42.8%; Pred. No. 2.9e-100;
Matches 390; Conservative 125; Mismatches 236; Indels 101; Gaps 27;

Qy 5 RGPLASLL--LLQVCMQCAASEPCRAVFEAEVTLBAGAEPPQALQKV-FMGC---58
Db 8 RGTLLPLLAALLQASVEASGEIALCKTGPEE-DVYSAVLPKDVHEGQPLLVKVFSCNRK 66

Qy 59 -----PQEPALFETDND--FTVRN-----GETVQER-----84
Db 67 RKQYSESEPADFKVDSGDTGVYAFPLTAEQAKFLIYAQDKETQEKQVAVNLGREPT 126

Qy 85 ---RSLKERNPLK--IPFSK-----RILRRHKRDVWVAPISVPENGKGPFPQRLNQLKSN 134
Db 127 LTBEPMPKEPHEIEBIVPRLAKHSGALQKQKDWIPPLNLPENSGRGPFPQELVIRSD 196

Qy 135 KDRTKFLPYITGQADSPGEGVFAVEKETGWLKLPDLREIAKYELPCHAVSNGAS 194
Db 187 RDKVLSRYSVTSGADQPPGTGFIINPISGQLSVTKPLDREIAFHFAHVAVDINGNQ 246

Qy 195 VEDPMNISIIITDQNDHKPKFTQDTFRGSLVGLPCTSVQVQATDDEDAIYTVNGVA 254
Db 247 VENPIDIVINDNDNRPEFLHQVNGSVPESSKFTYVMTATADDDP-NALNGMLR 305

Qy 255 YSHSQSPKPDHLMFTIHRSTGTISVSSGLDRKVPVETLTIQATMDGD---GSTTT 311
Db 306 YRILSQAPSTPSPNMFTINNETGDIITVAAGLDREKVVQYTLIIQATDMEGNPTYGLSNT 365

Qy 312 AVAVVEILDANDAPMFDPKQYEAHVPENAVGHVQELVTVDLDAPNSPAWRATYLMGG 371
Db 366 ATAVITVDNDNPETAMTFYGEVEPENVDIVANLVITDKDQPHTPAWNAAYISGG 425

Qy 372 DDGDHETIITHPESNOGILITTRKGLDPEAKNOHTLYVEVTNEAPFVLIK---PTSTATIV 428
Db 426 DPTGRFAITDPSNDGLVTVKPIDFETNRMFVLTVAAENQVPLAKG:QHPPESTATVS 485

Qy 429 VHVEDVNEAPFVFPSPKVVVEQEGIPTEGPFVVCVYTAEDPK-ENQKISYRIILRDPAGWLA 487
Db 486 VTVIDVNEPFAFNFPIIRQEBGLHAGTMLITLTAQDPDRYMQNIRYTKLSDPANWLK 545

Qy 488 MDPSQVTVAGTLDREDEQFVRNNIYFVWVLAMDNQSPPTGTGTLTLLTLDVNDHGPV 547
Db 546 IDPVNGQITIAVLDRS-SPYVQNNIYNATFLASDNQIPPMGSGTGTQLQIYLLDINDNAPQ 604

Qy 548 PEPRQITICNQSPVRHVLNIT--DKDLSPTGPFQALTRDSDIY---WTAEVNE-EGDT 601
Db 605 VLPQEAETC-ETPEPNSINIAALDYIDNAGFFAFDLPLSPVTIKGNWT--INRLNGDF 661

Qy 602 VVLSLK-KFLKQDTYDVHLSLDHGN--KEQUTVIRATVCDCHGH-----VETCPGPWK 653
Db 662 AQLNLKIKLEAGIYEVPIIITDSGNPKSNTSILRVKVCQDSNGDCTDVRIVGAGLG 721

Qy 654 GFTLPVLGAVLALLFLLVLLLV-----SKRKIKPEPLLPEDDTDNDNVFYGE 703
Db 722 -----TGAIILLCIIILLILVLMFVVMKRRKROAKQLIDPDVDDVEDNLIKYDE 775

Qy 704 EGGEBDDQYDITQHLRGLEAPV-----VLNDVAPTIIPTMYRPPA--NPEDEIGN 756
Db 776 EGGEBDDQYDLSQLQQDPTVEPDAIKPVGIRRLDERP-IHAEPQYVPSAAPHGPDIGD 834

Qy 757 FIENLKAANTDPTAPPYDTLIVDEYEGSGDAASLSLTSSASDQDQDYVLYNENGSPF 816
Db 835 FINEGKAADNDPTAPPYDLSLLVFDYEGSGTAGSLSLNSSSSGGDQDYDYLNDWGPFR 894

Qy 817 KKLADMYGGED 828
Db 895 KKLADMYGGDD 906

RESULT 15
IUCHCR
R-cadherin precursor - chicken
C/Species: Gallus gallus (chicken)
C/Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 16-Jun-2000
C/Accession: JH0424
R: Imazuka, H.; Miyatani, S.; Takeichi, M.
Neuron 7, 69-79, 1991
A/Title: R-cadherin: a novel Ca2+-dependent cell-cell adhesion molecule expressed in the
A/Reference number: JH0424; MUID:91299341; PMID:1712604
A/Accession: JH0424
A/Molecule type: mRNA
A/Residues: 1-913 <INU>
A/Cross-references: GB:D14459; GB:D00849; NID:g222854; PIDN:BAA03356.1; PID:g222855
A/Experimental source: retina
C/Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought to be involved in cell-cell adhesion.
C/Superfamily: cadherin; cadherin repeat homology
C/Keywords: calcium binding; cell adhesion; duplication; glycoprotein; retina; transmembrane
F:1-26/Domain: signal sequence #status predicted <SIG>
F:27-166/Domain: propeptide #status predicted <PRO>
F:167-913/Product: R-cadherin #status predicted <MAT>
F:167-721/Domain: extracellular #status predicted <EXT>
F:169-274/Domain: cadherin repeat homology <CR1>
F:244-249/Region: cadherin binding #status predicted
F:277-389/Domain: cadherin repeat homology <CR2>
F:392-504/Domain: cadherin repeat homology <CR3>
F:507-612/Domain: cadherin repeat homology <CR4>
F:613-721/Domain: transmembrane #status predicted <TM>
F:722-753/Domain: transmembrane #status predicted <TM>
F:754-913/Domain: intracellular #status predicted <INT>
F:870-885/Region: serine-rich
F:280,409,554,629,658,699/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 38.8%; Score 1694; DB 1; Length 913;
Best Local Similarity 41.5%; Pred. No. 5e-100;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 22:02:13 ; Search time 15 Seconds

(without alignments)

2877.744 Million cell updates/sec

Title: US-09-916-849A-1

Perfect score: 4369

Sequence: 1 MGLPRGPLASLLLLQVCWLQ.....NEWGRFVKLLADMYGGEDD 829

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt.42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	4369	100.0	829	CAD3_HUMAN	P22223 homo sapien
2	3550.5	81.3	822	CAD3_MOUSE	P10287 mus musculus
3	2573.5	58.9	732	CAD1_CHICK	P33145 gallus gall
4	2420	55.4	882	CAD1_HUMAN	P12830 homo sapien
5	2388	54.7	884	CAD1_MOUSE	P09803 mus musculus
6	2388	54.7	886	CAD1_RAT	Q90704 rattus norv
7	2375	54.3	880	CAD1_XENLA	P33148 xenopus lae
8	2371	54.3	884	CAD2_XENLA	P33152 xenopus lae
9	2323	53.2	491	CAD3_BOVIN	P19335 bos taurus
10	2274	52.0	887	CAD1_CHICK	P08841 gallus gall
11	2198	50.3	872	CAD1_XENLA	P30944 xenopus lae
12	1714	39.2	906	CAD2_HUMAN	P19022 homo sapien
13	1704.5	39.0	912	CAD2_CHICK	P10288 gallus gall
14	1697.5	38.9	906	CAD2_MOUSE	P15116 mus musculus
15	1694	38.8	913	CAD4_CHICK	P24503 gallus gall
16	1693	38.8	906	CAD2_RAT	Q92173 rattus norv
17	1691	38.7	877	CAD2_BOVIN	P19534 bos taurus
18	1681	38.5	893	CAD2_BRARE	Q90275 brachydanio
19	1655	37.9	913	CAD4_MOUSE	P39038 mus musculus
20	1621.5	37.1	916	CAD4_HUMAN	P55283 homo sapien
21	1616.5	37.0	906	CAD0_XENLA	P33147 xenopus lae
22	1609	36.8	905	CADN_XENLA	P20310 xenopus lae
23	1353.5	31.0	784	CAD2_MOUSE	P33446 mus musculus
24	1321.5	30.2	814	CAD2_HUMAN	P55291 homo sapien
25	1030.5	23.6	785	CAD7_HUMAN	Q9ulb5 homo sapien
26	1026	23.5	714	CADD_MOUSE	Q9wtr5 mus musculus
27	1021.5	23.4	713	CADD_HUMAN	P55290 homo sapien
28	1014	23.2	902	DSC2_MOUSE	P55292 mus musculus
29	1007.5	23.1	785	CAD7_CHICK	Q90763 gallus gall
30	990	22.7	790	CAD6_CHICK	Q90762 gallus gall
31	983.5	22.5	792	CAD6_CHICK	O93319 gallus gall
32	982	22.5	780	CAD6_HUMAN	P55285 homo sapien
33	976.5	22.4	790	CADI_HUMAN	Q13634 homo sapien

34	976	22.3	712	1	CADD_CHICK	P33150 gallus gall
35	975	22.3	789	1	CAD9_HUMAN	Q9ulb4 homo sapien
36	974	22.3	789	1	CAD6_RAT	P55280 rattus norv
37	969.5	22.2	895	1	DSC3_MOUSE	P55280 mus musculus
38	966.5	22.1	796	1	CAD2_HUMAN	P55287 homo sapien
39	963	22.0	863	1	DSC2_BOVIN	P33245 bos taurus
40	958.5	21.9	790	1	CAD6_MOUSE	P97326 mus musculus
41	957	21.9	896	1	DSC3_HUMAN	Q14574 homo sapien
42	951.5	21.8	796	1	CAD8_MOUSE	P55288 mus musculus
43	950	21.7	788	1	CADA_HUMAN	Q9Y6n8 homo sapien
44	948.5	21.7	901	1	DSC2_HUMAN	Q02487 homo sapien
45	940.5	21.5	896	1	DSC3_BOVIN	Q28060 bos taurus

ALIGNMENTS

RESULT 1
CAD3_HUMAN
ID_CAD3_HUMAN STANDARD; PRT; 829 AA.
AC P22223;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cadherin-3 precursor (Placental-cadherin) (P-cadherin).
GN CDH3 OR CDHP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90009051; PubMed=2793940;
RA Shimoyama Y., Yoshida T., Terada M., Shimosato Y., Abe O.,
RA Hirohashi S.;
RT "Molecular cloning of a human Ca2+-dependent cell-cell adhesion
RT molecule homologous to mouse placental cadherin: its low expression
RT in human placental tissues.";
RL J. Cell Biol. 109:1787-1794(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Strausberg R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Klausner R.D., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Pahey J., Helton E., Kettelman M., Madañ A., Rodrigues S., Sanchez A.,
RA Whiting M., Madañ A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Schmerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP DISEASE.
RX MEDLINE=21470328; PubMed=11544476;
RA Sprecher E., Bergman R., Richard G., Lurie R., Shalev S.,
RA Petronius D., Shalata A., Ambinder Y., Leibur R., Perlman I., Cohen N.,
RA Szargel R.;
RT "Hypotrichosis with juvenile macular dystrophy is caused by a
RT mutation in CDH3, encoding P-cadherin.";
RL Nat. Genet. 29:134-136(2001).
RN [4]

RP VARIANT HJMD HIS-503.
 RX MEDLINE=22336734; PubMed=12445216;
 RA Indelman M., Bergman R., Lurie R., Richard G., Miller B.,
 RA Petronius D., Ciubutaro D., Leib R., Sprecher E., causes
 RT "A missense mutation in CDH3, encoding p-cadherin, causes
 RL hypotrichosis with juvenile macular dystrophy.";
 RL J. Invest. Dermatol. 119:1210-1213(2002).
 CC -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
 CC They preferentially interact with themselves in a homophilic
 CC manner in connecting cells; cadherins may thus contribute to the
 CC sorting of heterogeneous cell types.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- DISEASE: Defects in CDH3 are the cause of hypotrichosis with
 CC juvenile macular dystrophy (HJMD) [MIM:601553]. HJMD is a rare
 CC autosomal recessive disorder characterized by early hair loss
 CC heralding severe degenerative changes of the retinal macula and
 CC culminating in blindness during the second to third decade of
 CC life.
 CC -!- SIMILARITY: Contains 5 cadherin domains.
 CC -!- DATABASE: NAME=Mutations of the CDH3 gene;
 CC NOTE=Retina International's Scientific Newsletter;
 CC WWW="http://www.retina-international.com/sci-news/cdh3mut.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X63629; CAA45177.1; -;
 DR EMBL; BC041846; AAH41846.1; -;
 DR PIR; A33659; IJHUCP.
 DR HSSP; P09803; 1SUH.
 DR Genew; HGNC:1762; CDH3.
 DR MIM; 114021; -;
 DR MIM; 601553; -;
 DR GO; GO:0007155; P:cell adhesion; TAS.
 DR InterPro; IPR002126; Cadherin.
 DR InterPro; IPR000233; Cadherin_C_term.
 DR Pfam; PF00028; cadherin; 5.
 DR Pfam; PF01049; Cadherin_C term; 1.
 DR PRINTS; PR00205; CADHERIN.
 DR SMART; SM00112; CA; 4.
 DR PROSITE; PS00232; CADHERIN_1; 3.
 DR PROSITE; PS0268; CADHERIN_2; 5.
 DR Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
 KW Signal; Disease mutation; Vision.
 FT SIGNAL 1 24
 FT PROPEP 25 107
 FT CHAIN 108 829
 FT DOMAIN 108 829 CADHERIN-3.
 FT TRANSMEM 108 854 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 655 677 POTENTIAL.
 FT DOMAIN 678 829 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 108 215 CADHERIN 1.
 FT DOMAIN 216 328 CADHERIN 2.
 FT DOMAIN 329 440 CADHERIN 3.
 FT DOMAIN 441 546 CADHERIN 4.
 FT DOMAIN 547 650 CADHERIN 5.
 FT DOMAIN 785 800 SER-RICH.
 FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 566 566 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 503 503 R -> H (in HJMD).
 FT /FTID=VAR_015422.
 FT CONFLICT 237 237 V -> M (IN REF. 2).
 FT CONFLICT 563 563 H -> Q (IN REF. 2).
 FT CONFLICT 829 829 AA; 91427 MW; ES03CFE5D981FI CRC64;
 SQ SEQUENCE 829 AA; 91427 MW; ES03CFE5D981FI CRC64;
 Query Match 100.0%; Score 4369; DB 1; Length 829;
 Best Local Similarity 100.0%; Pred. No. 2.7e-264;
 Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 MEDLINE=88111554; PubMed=3428270;

QY 1 MGLPRGPLASLLLLQVLCVWLOCAASEPCRAVPREAEVTLIAGGABQERGOALGVKVFMCSPG 60
 DB 1 MGLPRGPLASLLLLQVLCVWLOCAASEPCRAVPREAEVTLIAGGABQERGOALGVKVFMCSPG 60
 QY 61 CEPALFSTDNDDFTVRNGETQERSLKERNPLKIFPSKILRRHKRDVWVAPISVPENG 120
 DB 61 CEPALFSTDNDDFTVRNGETQERSLKERNPLKIFPSKILRRHKRDVWVAPISVPENG 120
 QY 121 KGPFFQRLNQLKSNKDRDKIFYSITGPGADSPPEGVFAVEKETGWLKLLNKLPLDREIAK 180
 DB 121 KGPFFQRLNQLKSNKDRDKIFYSITGPGADSPPEGVFAVEKETGWLKLLNKLPLDREIAK 180
 QY 181 YELFGHAYSENGASVEDPMNLSIIVTDQNDHKPFTQDTRGSLVLEGVLPQTSVMQVTAT 240
 DB 181 YELFGHAYSENGASVEDPMNLSIIVTDQNDHKPFTQDTRGSLVLEGVLPQTSVMQVTAT 240
 QY 241 DEDDAIYTYNGVWVAYS IHSQSPKDPHDLMTIHRSTGTISVISGLDREKVPYTLTQA 300
 DB 241 DEDDAIYTYNGVWVAYS IHSQSPKDPHDLMTIHRSTGTISVISGLDREKVPYTLTQA 300
 QY 301 TMDGSGSTTTAVAVVEILDANDNAPMDPKQYEAHVPENAVGHVQRLTVDLDAENSP 360
 DB 301 TMDGSGSTTTAVAVVEILDANDNAPMDPKQYEAHVPENAVGHVQRLTVDLDAENSP 360
 QY 361 AWRATYLLMGDDGDGHFTITTHPSNQGILTRKGLDFEAKNQHTLYVEVTNEAPFVLKL 420
 DB 361 AWRATYLLMGDDGDGHFTITTHPSNQGILTRKGLDFEAKNQHTLYVEVTNEAPFVLKL 420
 QY 421 PTSTATIIVHVEDVNEAPVFPVPPSKVQVQIGIPTGEPVCYTAEDDPKQKISYILR 480
 DB 421 PTSTATIIVHVEDVNEAPVFPVPPSKVQVQIGIPTGEPVCYTAEDDPKQKISYILR 480
 QY 481 DPAGWLAMDPPSGQVAVTGLDREDEQVRNNIYEVWVLMNDGSPPTTGTLTLLTLD 540
 DB 481 DPAGWLAMDPPSGQVAVTGLDREDEQVRNNIYEVWVLMNDGSPPTTGTLTLLTLD 540
 QY 541 VNDHGPVPEPRQITTCNQSPVRHVLNITDKLSPTSPFOQLTDDSDIYWTAEVNEGD 600
 DB 541 VNDHGPVPEPRQITTCNQSPVRHVLNITDKLSPTSPFOQLTDDSDIYWTAEVNEGD 600
 QY 601 TVVLSIKKFLKODTYDVHLSLSDHGNKQLTVIRATVCDCHGHVETCPGPKGFIPLVL 660
 DB 601 TVVLSIKKFLKODTYDVHLSLSDHGNKQLTVIRATVCDCHGHVETCPGPKGFIPLVL 660
 QY 661 GAVLALLFLLVLLVLLVRRKKIKLEPLLPLEDDTRDNVFFYGGEGGEDDQDITQLHR 720
 DB 661 GAVLALLFLLVLLVLLVRRKKIKLEPLLPLEDDTRDNVFFYGGEGGEDDQDITQLHR 720
 QY 721 GLEAPVWVLRNDVAPTIPTPMYRPRANDEIGNFIENLKAANTDPTAPPYDTLLVF 780
 DB 721 GLEAPVWVLRNDVAPTIPTPMYRPRANDEIGNFIENLKAANTDPTAPPYDTLLVF 780
 QY 781 DYEGSGSDAASLSLTSASDQDQDYLYNEWGSRFKKLADMYGGEDD 829
 DB 781 DYEGSGSDAASLSLTSASDQDQDYLYNEWGSRFKKLADMYGGEDD 829

RESULT 2

CAD3_MOUSE
 ID_CAD3_MOUSE STANDARD; PRT; 822 AA.
 AC P10387; Q61465;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cadherin-3 precursor (Placental-cadherin) (p-cadherin).
 GN CDH3 OR CDHP.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88111554; PubMed=3428270;

RA Nose A., Nagafuchi A., Takeichi M.;
RT "Isolation of placental cadherin cDNA: identification of a novel gene
family of cell-cell adhesion molecules.;"
RL EMO J. 6:3655-3661(1987).
RN [2]
RP PARTIAL SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Embryo;
RX MEDLINE=91360343; PubMed=1886768;
RA Hattori M., Miyatani S., Copeland N.G., Gilbert D.J., Jenkins N.A.,
RA Takeichi M.;
RT "Genomic organization and chromosomal mapping of the mouse P-cadherin
gene.;"
RL Nucleic Acids Res. 19:4437-4441(1991).
RN [3]
RP SEQUENCE OF 1-55 FROM N.A.
RC STRAIN=C57BL/6;
RX STRAIN=C57BL/6;
RA Hattori M., Takeichi M.;
RT "Complex cell type-specific transcriptional regulation by the promoter
and an intron of the mouse P-cadherin gene.;"
RL Dev. Growth Differ. 36:509-519(1994).
RN [4]
RP DEVELOPMENTAL STAGE.
RC STRAIN=C57BL/6; TISSUE=Testis;
RX MEDLINE=97033837; PubMed=8879495;
RA Munro S.B., Blaschuk O.W.;
RT "A comprehensive survey of the cadherins expressed in the testes of
fetal, immature, and adult mice utilizing the polymerase chain
reaction.;"
RL Biol. Reprod. 55:822-827(1996).
RN [5]
RP FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
CC They preferentially interact with themselves in a homophilic
manner in connecting cells; cadherins may thus contribute to the
sorting of heterogeneous cell types.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DEVELOPMENTAL STAGE: Expression is high in both fetal and newborn
testis but minimal in testis of 7-day-old animals. Not detected in
testis of 21-day-old or adult.
CC -1- SIMILARITY: Contains 5 cadherin domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X06340; CAA29646.1; -;
DR EMBL; D12688; BAA02186.1; -;
DR PIR; S03163; IJMSUP.
DR HSP; P09803; IJMSUP.
DR MGD; MGI:88356; Cdh3.
DR GO; GO:0005886; C:plasma membrane; IDA.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR000233; Cadherin_C_term.
DR Pfam; PF00028; cadherin; 5.
DR Pfam; PF01049; Cadherin C term; 1.
DR PRINTS; PR00205; CADHERIN.
DR SMART; SM00112; CA; 4.
DR PROSITE; PS00232; CADHERIN_1; 3.
DR PROSITE; PS00268; CADHERIN_2; 4.
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
Signal.
FT SIGNAL 1 25 POTENTIAL.
FT PROPEP 26 99
FT CHAIN 100 822
FT DOMAIN 100 647
FT TRANSMEM 648 670
FT DOMAIN 671 822
FT DOMAIN 823 827
FT DOMAIN 828 832
FT DOMAIN 833 837
FT DOMAIN 838 842
FT DOMAIN 843 847
FT DOMAIN 848 852

FT DOMAIN 539 645 CADHERIN 5.
FT DOMAIN 773 793 SER-RICH.
FT CARBOHYD 192 192 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 558 558 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 822 AA; 90754 MW; 84975D2115B14DC3 CRC64;

Query Match 81.3%; Score 3550.5; DB 1; Length 822;
Best Local Similarity 81.5%; Pred. No. 2.3e-213;
Matches 678; Conservative 57; Mismatches 84; Indels 13; Gaps 5;

QY 1 MGLPRGLA-SLLILQVWLCQAASEPCRAVF-REAEVTLKAGAEQEPGQALGVKFWGC 58
DB 1 MELLSPHAFLLLLQVWLCRSVSEPYRAGFIGEAGVTLEVEGTDLEPSQVLGVKVALAG 60

QY 59 PQCEPALFSTDDFTVANGETVQERRSLKERNPLKIPFSKRILRRHRDWWVAPISVPE 118
DB 61 QCVHHA---DNGDIIMLTRGTVOGKDAMES-----PPTILRRRRREWWMPPIFVPE 110

QY 119 NGKGPFPORLNQKSKORDTKIFYSITGPGADSPPEGVFAVEKGTGMLLNKPLDREI 178
DB 111 NGKGPFPORLNQKSKDRGTFKIFYSITGPGADSPPEGVFTIEKSGWLLHMPDREKI 170

QY 179 AKVELFGHAVSENGASVEDPMNISIIIVTDQNDHKFTQDTRFGSVLGLVPGTSMQVT 238
DB 171 VKELYGHAVSENGASVEEPMNISIIIVTDQNDKFKFTQDTRFGSVLGLVPGTSMQVT 230

QY 239 ATDEDAIYTYNGVAVSIHQEPKPHDLMPHTIHRSTGTISVTSSGLDREKVPYTLTI 298
DB 231 ATDEDAVNTYNGVAVSIHQEPKPHDLMPHTIHKSTGTISVTSSGLDREKVPYRLTV 290

QY 299 QATMDGSGSTTAVAVVILLDANNAWFOPOKYEAAHVAVENAVGHEVQRTVTDLDPN 358
DB 291 QATMDGSGSTTAVAVVILLDANNAWFOPEPQKYEAAVAVENAVGHEVQRTVTDLDPN 350

QY 359 SPAMRATYILMGDDGDDHFTITTHPESQGLITTRKGLDFAKNQHTLYVEVNEAPVL 418
DB 351 WPAWRATYIHVGGDDGDDHFTITTHPESQGLITTRKGLDFAKNQHTLYVEVNEAPVL 410

QY 419 KLPTATYIVHVEVNEAPVFPVPSKVVEQEGTPTCEPVCVYTAEPDKNQKISRI 478
DB 411 KLPTATYIVHVEVNEAPVFPVPSKVVEQEGTPTCEPVCVYTAEPDKNQKISRI 470

QY 479 LDPAGLAMPDSQVAVGLTDRDEQFVNNIYEVWVWVWVWVWVWVWVWVWVWVWVWVW 538
DB 471 SEDPANLAVPDSQVAVGLTDRDEQFVNNIYEVWVWVWVWVWVWVWVWVWVWVWVW 530

QY 539 IDVDHGVPEPRQITICNQSPVRHLNITDKDLSPTSPFOAQLTDDSDIYWTAEVNEE 598
DB 531 TDINDHGPTEPRQITICNQSPVRHLNITDKDLSPTSPFOAQLTDDSDIYWTAEVNEE 590

QY 599 GDTVVLKFKLQDQYDVHLSDPHGKNEQLTVIRATVCDCHGV-ETCPGPKWGGFLL 657
DB 591 GDTVVLKFKLQDQYDVHLSDPHGKNEQLTVIRATVCDCHGV-ETCPGPKWGGFLL 650

QY 658 PVLGAVLALLFLLVLLVLRKKRKKIKPEPLLPEDDTRDNVFFYEGEGGEDQYDITQ 717
DB 651 PVLGAVLALLFLLVLLVLRKKRKKIKPEPLLPEDDTRDNVFFYEGEGGEDQYDITQ 710

QY 718 LHRGLEAPVVLNDVAPTITPTMYRPRANDEIGNFIENLKANTPTAPPYDTL 777
DB 711 LHRGLEAPVVLNDVAPTITPTMYRPRANDEIGNFIENLKANTPTAPPYDTL 770

QY 778 LVFDYEGSGDAASLSLTSASDQDQYDYLNEGWSRFFKKLADMYGGEDD 829
DB 771 LVFDYEGSGDAASLSLTSASDQDQYDYLNEGWSRFFKKLADMYGGEDD 822

RESULT 3
ID CADL CHICK STANDARD; PRT; 732 AA.
AC P33175;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE B-cadherin precursor (K-CAM protein) (Fragment).
GN K-CAM.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92107987; PubMed=1763068;
RA "Sorkin B.C., Gallin W.J., Edelman G.M., Cunningham B.A.;
RT "Genes for two calcium-dependent cell adhesion molecules have similar
RT structures and are arranged in tandem in the chicken genome.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:11545-11549(1991).
RN [2]
RP SEQUENCE OF 7-732 FROM N.A.
RC TISSUE=Embryonic brain;
RX MEDLINE=91225083; PubMed=2026653;
RA Napolitano E.W., Venstrom K., Wheeler E.F., Reichardt L.F.;
RT "Molecular cloning and characterization of B-cadherin, a novel chick
RT cadherin.";
RL J. Cell Biol. 113:893-905(1991).
CC -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
CC They preferentially interact with themselves in a homophilic
CC manner in connecting cells; cadherins may thus contribute to the
CC sorting of heterogeneous cell types. B-cadherin may have important
CC functions in neurogenesis, in at least some epithelia, and in
CC embryogenesis.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: Expressed in a wide variety of tissues.
CC -!- SIMILARITY: Contains 5 cadherin domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M81894; AAA48929.1; -;
DR EMBL; X58518; CA41408.1; -;
DR PIR; A41634; IJCHCB.
DR HSSP; P09803; 1SUH.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR002233; Cadherin_C_term.
DR Pfam; PF00028; cadherin; 5.
DR Pfam; PF01049; Cadherin_C_term; 1.
DR PRINTS; PR00205; CADHERIN.
DR PROSITE; PS00232; CADHERIN_1; 3.
DR PROSITE; PS0268; CADHERIN_2; 5.
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
FT NON TER 1 1
FT PROPEP 1 6
FT CHAIN 7 732
FT DOMAIN 6 554
FT TRANSMEM 555 580
FT DOMAIN 581 732
FT CYTOPLASMIC (POTENTIAL).
FT DOMAIN 1.
FT CADHERIN 1.
FT DOMAIN 2.
FT CADHERIN 2.
FT DOMAIN 3.
FT CADHERIN 3.
FT DOMAIN 4.
FT CADHERIN 4.
FT DOMAIN 5.
FT CADHERIN 5.
FT SER-RICH.
FT DOMAIN 689 702
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 137 137
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 410 410
FT M -> V (IN REF. 2).
FT CONFLICT 414 414
FT SEQUENCE 732 AA; 80613 MW; 091D59A6A16CBD45 CRC64;

Query Match 58.9%; Score 2573.5; DB 1; Length 732;
Best Local Similarity 66.2%; Pred. No. 1.3e-152;
Matches 485; Conservative 103; Mismatches 138; Indels 7; Gaps 5;

QY 102 LRRHKRWVAPISVPEKGGPFQRLNQLKSNKDRDTKIFYSITGPAGSDPPPEGVAVE 161
DB 1 LRRQRQWVPIPKVPENERGFFPNLVQIAKSNDRRAKIFYSITGGADAPPISGITE 60
QY 162 KETGMLLNKLPDREELIAKELFCHAVSENGASVEDPMN:SIIVTDQNDHKPKETQTFR 221
DB 61 KETGMKVTQPLDREHINKYHLYSHAVSENGKPEEPMEIIVTVDQNDKPKQTOEVR 120
QY 222 GSVLEGVLPCTSVNQVATDDEDAIYNGVAVYSIHSQEPKDPHDLMEFTHRSTGISV 281
DB 121 GSVPEGALPGTSVMRVNATDADDVETVNGVIAYSIQSPEPHPHMFTVNRATGLSV 180
QY 282 ISSGLDRKVEPYLTITQATDMGDGSGTTTAVAVVEILDANDNAPMFDPPQYEAHVENA 341
DB 181 IASGLDRERVREYLTJMQAADLGGLTTTALAIVETDVNDNAPEFDPKTYEAAVENE 240
QY 342 VGHEVQRLTVTDLAPNPAWRATYILMGDDGDGHFTITTHPESNOGILLTRKGLDEAK 401
DB 241 AELEVARLATTDLDEPHTPAWRAVYSIVRGEGGAFTITDPASNEGVLRKAGLDEAK 300
QY 402 NCHTLVYVETNEAPFVLKLPSTATIVVHVEDVNEAPVFVPPSKVVEQEGIPTEGPVCV 461
DB 301 RQFVHLVAVNEAPPAIKLPATATVMVSEDVNEAPVDFPPLAQVPEDVPLGQPLAS 360
QY 462 YTAEDPK-ENQKISYRILRDPAGWLAMPDSGQVAVGTLDREDEOFVRNIIYEVWLA 520
DB 361 YTAQDPDRAQQRIKYVMGSDPAGWLAVHPENGLITAREQLDRE-SPFTKNSTNYAVLLA 419
QY 521 MDNGSPPTGTGTLTLTLDVNDHGPVPEPQITICNQSPVRHYLVNITDKDLSHTSPFO 580
DB 420 VDDGPPATGTLTLTLDVNDHGPVPEPQITICNQSPVRHYLVNITDKDLSHTSPFO 479
QY 581 AQLTDDSDIYNTAEVNEEGDTVLSLKKFLKODTYDVHLSLSDHGNKEQLTVIRATVCDC 640
DB 480 AELSHGSGDSWAVEGVNGGDTVALWLTPELQNLVSVVRLFDQKGDQVTVIRAQVDC 539
QY 641 HGVETC---PGPWKG-CFILPVIGAVLAFLVLLVLLVLRKKIKPELPLLEDTRD 696
DB 540 QQRVESCAQKPRVDTGVFVILVAVLAVLALLVLLVLLVLRKKIKPELPLLEDTRD 599
QY 697 NVFYEGEGGGEEDQYDITQLHRLGLARPVLNVDVAPTIPTPMYRPPANPDIGN 756
DB 600 NIFYEGEGGGEEDQYDLSQLHRLGLARPEVI-RNDVAPPLMAAPQYRPPANPDIGN 658
QY 757 FIENLKAANTDPTAPPYDITLLVFDYEGSGDAASLSLTSASDODDYVLEWNSRF 816
DB 659 FIDENLKAADTPTAPPYDITLLVFDYEGSGSEATSLSLNSSASDQDDQDYDLNENWGRF 718
QY 817 KKLADMYGGGEDD 829
DB 719 KKLAELYGGGEDE 731

RESULT 4
CAD1_HUMAN STANDARD; PRT: 882 AA.
ID_CAD1_HUMAN Q13799; Q14216; Q15855; Q16194;
AC P12870; Q13799; Q14216; Q15855; Q16194;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Epithelial-cadherin precursor (E-cadherin) (Uvomorulin) (cadherin-1)
DE (CAM 120/80).
GN CDH1 OR UVO OR CDHE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=93211394; PubMed=8459805;
RA Bussemakers M.J.G., Mees S.G.M., van Bokhoven A., Debruyne F.M.J.,

RA Schalken J.A.;
RT "Molecular cloning and characterization of the human E-cadherin
RL cDNA.";
RN Mol. Biol. Rep. 17:123-128(1993).
[2]
RP SEQUENCE FROM N.A.
RA Kelker W., Warda A., Oda T., Hirohashi S., Kemler R., Birchmeier W.;
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=94242050; PubMed=8185635;
RA Rimm D.L., Morrow J.S.;
RT "Molecular cloning of human E-cadherin suggests a novel subdivision of
RL the cadherin superfamily.";
RN Biochem. Biophys. Res. Commun. 200:1754-1761(1994).
[4]
RP SEQUENCE OF 56-882 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=95324920; PubMed=7601454;
RA Bex G., Staes K., van Hengel J., Molemans F., Bussemakers M.J.G.,
van Bokhoven A., van Roy F.;
RT "Cloning and characterization of the human invasion suppressor gene
RL E-cadherin (CDH1).";
RN Genomics 26:281-289(1995).
[5]
RP SEQUENCE OF 172-311 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=89031725; PubMed=3263290;
RA Mansouri A., Spurr N., Goodfellow P.N., Kemler R.;
RT "Characterization and chromosomal localization of the gene encoding
RL the human cell adhesion molecule uvomorulin.";
RN Differentiation 38:67-71(1988).
[6]
RP SEQUENCE OF 265-392 FROM N.A.
RC TISSUE=Liver;
RA Frixen U.H.;
RL Submitted (MAR-1990) to the EMBL/GenBank/DBJ databases.
[7]
RP SEQUENCE OF 1-16 FROM N.A.
RX MEDLINE=94380041; PubMed=8093045;
RA Bussemakers M.J.G., Girolidi L.A., van Bokhoven A., Schalken J.A.;
RT "Transcriptional regulation of the human E-cadherin gene in human
RL prostate cancer cell lines: characterization of the human E-cadherin
RN gene promoter.";
RN Biochem. Biophys. Res. Commun. 203:1284-1290(1994).
[8]
RP SEQUENCE OF 1-16 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=95365379; PubMed=7543680;
RA Yoshiura K., Kanai Y., Ochiai A., Shimoyama Y., Sugimura T.,
Hirohashi S.;
RT "Silencing of the E-cadherin invasion-suppressor gene by CpG
RL methylation in human carcinomas";
RN Proc. Natl. Acad. Sci. U.S.A. 92:7416-7419(1995).
[9]
RP INTERCHAIN DISULFIDE BOND.
RX MEDLINE=21975237; PubMed=11856755;
RA Makagiansar I.T., Nguyen P.D., Ikesue A., Kucera K., Dentler W.,
Urbauer J.L., Galeva N., Alterman M., Sahaan T.J.;
RT "Disulfide bond formation promotes the cis- and trans-dimerization of
RL the E-cadherin-derived first repeat.";
RN J. Biol. Chem. 277:16002-16010(2002).
[10]
RP DISEASE.
RX MEDLINE=99406500; PubMed=10477433;
RA Guilford P.J., Hopkins J.B.W., Grady W.M., Markowitz S.D., Willis J.,
Lynch H., Rajput A., Wiesner G.L., Lindor N.M., Burgart L.J.,
Toro T.F., Lee D., Limacher J.-M., Shaw D.W., Findlay M.P.N.,
Reeve A.E.;
RT "E-cadherin germline mutations define an inherited cancer syndrome
RL dominated by diffuse gastric cancer.";
RN Hum. Mutat. 14:249-255(1999).
[11]
RP REVIEW ON VARIANTS.
RX MEDLINE=98415721; PubMed=9744472;
RA Bex G., Becker K.-F., Hoefler H., van Roy F.;
RT "Mutations of the human E-cadherin (CDH1) gene.";
RN Hum. Mutat. 12:226-237(1998).
[12]
RP SEQUENCE OF 337-476 FROM N.A., AND VARIANTS ALA-370 AND ASP-473.
RX MEDLINE=94306394; PubMed=8033105;
RA Becker K.-F., Atkinson M.J., Reich U., Becker I., Nekarda H.,
Siewert J.R., Hoefler H.;
RT "E-cadherin gene mutations provide clues to diffuse type gastric
RL carcinomas.";
RN Cancer Res. 54:3845-3852(1994).
[13]
RP VARIANT LOBULAR BREAST CARCINOMA SER-315.
RX MEDLINE=95049851; PubMed=7961105;
RA Kanai Y., Oda T., Teuda H., Ochiai A., Hirohashi S.;
RT "Point mutation of the E-cadherin gene in invasive lobular carcinoma
RL of the breast.";
RN Jpn. J. Cancer Res. 85:1035-1039(1994).
[14]
RP VARIANTS GYNECOLOGIC CANCERS THR-617; VAL-711 AND GLY-838.
RX MEDLINE=94355985; PubMed=8075649;
RA Risinger J.I., Berchuck A., Kohler M.F., Boyd J.;
RT "Mutations of the E-cadherin gene in human gynecologic cancers.";
RN Nat. Genet. 7:98-102(1994).
[15]
RP VARIANT GASTRIC ADENOCARCINOMA 274-GLY--PRO-277 DEL.
RX MEDLINE=94173928; PubMed=8127895;
RA Oda T., Kanai Y., Oyama T., Yoshiura K., Shimoyama Y., Birchmeier W.,
Sugimura T., Hirohashi S.;
RT "E-cadherin gene mutations in human gastric carcinoma cell lines.";
RN Proc. Natl. Acad. Sci. U.S.A. 91:1858-1862(1994).
[16]
RP VARIANT GASTRIC CARCINOMA PRO-193.
RX MEDLINE=96390918; PubMed=8797891;
RA Muta H., Noguchi M., Kanai Y., Ochiai A., Nawata H., Hirohashi S.;
RT "E-cadherin gene mutations in signet ring cell carcinoma of the
RL stomach.";
RN Jpn. J. Cancer Res. 87:843-848(1996).
[17]
RP VARIANTS GASTRIC CARCINOMA ASP-400 DEL AND 418-ASP--PHE-423 DEL.
RX MEDLINE=97197648; PubMed=9045944;
RA Tamura G., Sakata K., Nishizuka S., Maesawa C., Suzuki Y., Iwaya T.,
Terashima M., Saito K., Satodate R.;
RT "Inactivation of the E-cadherin gene in primary gastric carcinomas and
RL gastric carcinoma cell lines.";
RN Jpn. J. Cancer Res. 87:1153-1159(1996).
[18]
RP VARIANT THYROID TUMOR THR-592.
RX MEDLINE=97138061; PubMed=8985087;
RA Soares P., Bex G., van Roy F., Sobrinho-Simoes M.;
RT "E-cadherin gene alterations are rare events in thyroid tumors.";
RN Int. J. Cancer 70:32-38(1997).
[19]
RP VARIANTS ASP-336 AND ILE-470.
RX MEDLINE=98196671; PubMed=9537325;
RA Guilford P.J., Hopkins J.B.W., Harraway J., McLeod M., McLeod N.,
Harawita P., Taite H., Scoular R., Miller A., Reeve A.E.;
RT "E-cadherin germline mutations in familial gastric cancer.";
RN Nature 392:402-405(1998).
[20]
RP VARIANTS HDGC GLY-244 AND ALA-487.
RX MEDLINE=99253140; PubMed=10319582;
RA Yoon K.-A., Ku J.-L., Yang H.-K., Kim W.H., Park S.Y., Park J.-G.;
RT "Germline mutations of E-cadherin gene in Korean familial gastric
RL cancer patients.";
RN J. Hum. Genet. 44:177-180(1999).
[21]
RP VARIANT COLORECTAL CANCER ALA-340.
RX MEDLINE=20357134; PubMed=10896919;
RA Kim H.C., Wheeler J.M.D., Kim J.C., Ilyas M., Beck N.E., Kim B.S.,

RA Park K.C., Bodmer W.F.;
RT "The E-cadherin gene (CDH1) variants T340A and L599V in gastric and
RL colorectal cancer patients in Korea.";
RN Gut 47:262-267(2000).
RP
RX VARIANT ALA-270.
RX MEDLINE=21562463; PubMed=11705864;
RA Ikonen T., Matikainen M., Mononen N., Hyyinen E.R., Helin H.J.,
RA Tommila S., Tammela T.L., Pukkala E., Schleutker J., Kallioniemi O.P.,
RA Kivisto P.A.;
RT "Association of E-cadherin germ-line alterations with prostate
RT cancer.";
RL Clin. Cancer Res. 7:3465-3471(2001).
RN
RN VARIANT THR-592.
RX MEDLINE=21446527; PubMed=11562785;
RA Salahshor S., Hou H., Diep C.B., Loukola A., Zhang H., Liu T.,
RA Chen J., Iselius L., Rubio C., Lothe R.A., Aaltonen L., Sun X.F.,
RA Lindmark G., Lindblom A.;
RT "A germline E-cadherin mutation in a family with gastric and colon
RT cancer.";
RL Int. J. Mol. Med. 8:439-443(2001).
RN
RN VARIANT HDGC ALA-340.
RX MEDLINE=22021361; PubMed=11968083;
RA Oliveira C., Bordin M.C., Grehn N., Huntsman D., Suriano G.,
RA Machado J.C., Kivulovu T., Aaltonen L., Jackson C.E., Seruca R.,
RA Caldas C.;
RT "Screening E-cadherin in gastric cancer families reveals germline
RT mutations only in hereditary diffuse gastric cancer kindred.";
RL Hum. Mutat. 19:510-517(2002).
CC -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
CC They preferentially interact with themselves in a homophilic
CC manner in connecting cells; cadherins may thus contribute to the
CC sorting of heterogeneous cell types. E-cadherin has a potent
CC invasive suppressor role. It is also a ligand for integrin alpha-
CC E/beta-7.

Query Match 55.4%; Score 2420; DB 1; Length 882;
Best Local Similarity 55.1%; Pred. No. 5.8e-143; Gaps 13;
Matches 483; Conservative 112; Mismatches 222; Indels 60; Caps 13;
QY 9 ASLLQLQV-CWLOCASEPRAVREAEVLEAGAEQEPGQALGV-FMGCPG-QEPAL 65
DB 10 ALLLLQVSSL-CQEPCHPGDAESYFTVPRHLERGRVGRVNFEDCTGRTAY 68
QY 66 PSTNDPTVRNGTVERSLKERNP----- 92
DB 69 FSLDT-RFKVGTGVIITKGRPLRPHNPOIHFLVYVNDSTYRKFTSKVTLNVTGHHRRPP 127
QY 93 -----LKIFP-SKILRHKEDVWVAPISVPENKGPFPQRLNQLKNDKRDYK 140
DB 128 HQASVGIQAEILLTFPNSPCLRQKQDWIPIPCPENKGPFPKLVQIKSKDKGK 167
QY 141 IFYSITGPGADSPGEPFAVEKETGMLLNKPLDREBIAYELFCHAVSNGASVEDPMN 200
DB 188 VFYSITGQADTPGVGFIHERETGMLKTEPLDRERIATYTLFSAVSSNGNAVEDPM 247
QY 201 ISIVTQNDHKPFTDTRGSLVLEGVLPQTSVMQVATDDEDAIYTVNGVAVYSHQ 260
DB 248 ILIIVTQNDKPKFTQEVGKSVMEALPGTSVMEVATDADDVNTYNAATAYILSQ 307
QY 261 EPKDPHLMFTIHRSTGTISVSSGLDREKVPYETLTITQATMDGSGSTTAVAVVILD 320
DB 308 DPFLPDKNMFTINRTGIVSVVITGLDRESFPYTLVQQAADLQGEGLSTATAVITVD 367
QY 321 ANDNAPMFDPOKFAHVPENAVGVEQLTVTDLAPNSPAWRATYLMGGDGDHFTT 380
DB 368 TNDNPPFNPTYGQFNEANVITLKVTDADANTPAWEAVITIL-NDGGQFVVT 426
QY 381 THPESNOGILTRKGLDPEAKNQHTLVYVNTNEAPFVLKLPSTATVHVHEDVNEAPVF 440
DB 427 TNPVNDGILKTAAGLDFEAKQYILHVAVTNVVPFVSLTITATVTVDLVDVNEAPIF 486

QY 441 VPPSKVVEVQSGIPTGEPVVCVYTAEDDPK-ENQKISYRILRDPAGWLAMDPSQGVTAVG 499
DB 487 VPPEKRVESVEDFGVQGEITSYTAQEPDTEMEQKITVIRWDRDTANWLEINPDTGAISTRA 546
QY 500 TLDRDEQFVNNIYEVWVLAMDNGSPPTTGTGLLTLLIDVNDHGPVPPRQITICNQS 559
DB 547 ELDRDFPHVKNSTYATLIATDNGSPVATGTGILLIILSDVNDNAPIPRPTIFFCERN 606
QY 560 PVRHVINTKDLSEHTSPFOAQLTDSDIYWTAEVNE-EGDTVVLSELKFLKODTYDMH 618
DB 607 PKPQVINIIDADLPENTSPFTAEILTHGASANNTIYNDPTQESILKPKQMALEVGYKIN 666
QY 619 LLSLDHGNKEQLTVIRATVCDCHGHVETC--PGWKGKF-----LPLVGLAVLALLELLV 672
DB 667 LKLMNDQNKQVTTLEVSVCCEGAAGVCRKAQVEAGLQIPAILGILGGLALLIILL 726
QY 673 LLLLVRKXKIKIPELPLLPEDDTRDNVFFYEGEGGEDQDYDITQLHRGLEAREPVVLRN 732
DB 727 LLLFLRRRAVVKPELPPEDDTRDNVFFYEGEGGEDQDFDLSQLHRLDAREVT-RN 785
QY 733 DVAPTIIPMYRPRPANDPDEIGNFIENLKAANTDTPAPPYDILLVDFEGSGSDAASL 792
DB 786 DVAPTIIPMYRPRPANDPDEIGNFIENLKAANTDTPAPPYDILLVDFEGSGSDAASL 845
QY 793 SLSLTSSASDQDQDYDLNENWGRFCKLADMYGGGEDD 829
DB 846 SLSLTSSASDQDQDYDLNENWGRFCKLADMYGGGEDD 882

RESULT 5
CAD1 MOUSE STANDARD; PRT; 884 AA.
ID CAD1 MOUSE
AC P09603; Q61377;
DT 01-VAR-1989 (Rel. 10, Created)
DT 01-VAR-1989 (Rel. 10, last sequence update)
DT 15-VAR-2004 (Rel. 43, last annotation update)
DE Epithelial-cadherin precursor (E-cadherin) (Uvomorulin) (Cadherin-1)
DE (ARC-1).
GN CDH1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
ON [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR.
RX MEDLINE=87315445; PubMed=3498123;
RX Nagafuchi A., Shirayoshi Y., Okazaki K., Yasuda K., Takeichi M.;
RT "Transformation of cell adhesion properties by exogenously introduced
RT E-cadherin cDNA.";
RL Nature 329:341-343(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/Sv.
RX MEDLINE=92093614; PubMed=1754391;
RX Ringwald M., Baribault H., Schmidt C., Kemler R.;
RT "The structure of the gene coding for the mouse cell adhesion
RT molecule uvomorulin.";
RL Nucleic Acids Res. 19:6533-6539(1991).
RN [3]
RP SEQUENCE OF 174-884 FROM N.A. AND SEQUENCE OF 157-181.
RX MEDLINE=88111553; PubMed=3501370;
RX Ringwald M., Schun R., Vestweber D., Eistetter H., Lottspeich F.,
RX Engel J., Doeiz R., Jaehnic F., Epplen J., Mayer S., Mueller C.,
RX Kemler R.;
RT "The structure of cell adhesion molecule uvomorulin. Insights into
RT the molecular mechanism of Ca2+-dependent cell adhesion.";
RL EMBO J. 6:3647-3653(1987).
RN [4]
RP SEQUENCE OF 1-15 FROM N.A.
RX MEDLINE=92107977; PubMed=1763063;
RX Behrens J., Loewrick O., Klein-Hitpass L., Birchmeier W.;

RT "The E-cadherin promoter: functional analysis of a G.C-rich region
 RT and an epithelial cell-specific palindromic regulatory element.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:11495-11499(1991).
 RN [5]
 RP DEVELOPMENTAL STAGE.
 RC STRAIN=C57BL/6; TISSUE=Testis;
 RX MEDLINE=97033837; PubMed=8879495;
 RA Munro S.B., Blaschuk O.W.;
 RT "A comprehensive survey of the cadherins expressed in the testes of
 RT fetal, immature, and adult mice utilizing the polymerase chain
 RT reaction.";
 RL Biol. Reprod. 55:822-827(1996).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 157-370.
 RX MEDLINE=96176249; PubMed=8598933;
 RA Nagar B., Overduin M., Ikura M., Rini J.M.;
 RT "Structural basis of calcium-induced E-cadherin rigidification and
 RT dimerization.";
 RL Nature 380:360-364(1996).
 RN [7]
 RP STRUCTURE BY NMR OF 157-260.
 RX MEDLINE=96271285; PubMed=8785495;
 RA Overduin M., Tong X.J., Kay C.M., Ikura M.;
 RT "1H, 15N and 13C resonance assignments and monomeric structure of the
 RT amino-terminal extracellular domain of epithelial cadherin.";
 RL J. Biomol. NMR 7:173-189(1996).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS) OF 577-728 IN COMPLEX WITH
 RP CTNNB1, AND PHOSPHORYLATION.
 RX MEDLINE=21246507; PubMed=11348595;
 RA Huber A.H., Weiss W.I.;
 RT "The structure of the beta-catenin/E-cadherin complex and the
 RT molecular basis of diverse ligand recognition by beta-catenin.";
 RL Cell 105:391-402(2001).
 CC -1- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
 CC They preferentially interact with themselves in a homophilic
 CC manner in connecting cells; cadherins may thus contribute to the
 CC sorting of heterogeneous cell types. E-cadherin is a ligand for
 CC integrin alpha-E/beta-7.
 CC -1- SUBUNIT: Homodimer. Binds CTNNB1.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: Non-neutral epithelial tissues.
 CC -1- DEVELOPMENTAL STAGE: In the testis, expression is highest in fetal
 CC gonad, then decreases 5-fold in newborn. Detectable in 7-day-old
 CC but not in 21-day-old or adult.
 CC -1- SIMILARITY: Contains 5 cadherin domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X06115; CAA29488.1; -
 DR EMBL; X60961; CAA43292.1; -
 DR EMBL; X60962; CAA43292.1; JOINED.
 DR EMBL; X60963; CAA43292.1; JOINED.
 DR EMBL; X60964; CAA43292.1; JOINED.
 DR EMBL; X60965; CAA43292.1; JOINED.
 DR EMBL; X60966; CAA43292.1; JOINED.
 DR EMBL; X60967; CAA43292.1; JOINED.
 DR EMBL; X60968; CAA43292.1; JOINED.
 DR EMBL; X60969; CAA43292.1; JOINED.
 DR EMBL; X60970; CAA43292.1; JOINED.
 DR EMBL; X60971; CAA43292.1; JOINED.
 DR EMBL; X60972; CAA43292.1; JOINED.
 DR EMBL; X60973; CAA43292.1; JOINED.
 DR EMBL; X60974; CAA43292.1; JOINED.
 DR EMBL; X60975; CAA43292.1; JOINED.
 DR EMBL; X08339; CAA29645.1; -
 DR EMBL; M81449; AAA37352.1; -

DR PIR; S04528; IJMSCE.
 DR PIR; S34438; S34438.
 DR PDB; 1EDH; 11-JAN-97.
 DR PDB; 1SUH; 11-JUL-96.
 DR PDB; 117W; 09-MAY-01.
 DR PDB; 117X; 16-MAY-01.
 DR PDB; 1PF5; 23-AUG-00.
 DR MGD; MGI:88354; Cdh1.
 DR GO; GO:0005886; C:Plasma membrane; IDA.
 DR GO; GO:0005503; F:calcium ion binding; IDA.
 DR GO; GO:0042804; F:protein homooligomerization activity; IDA.
 DR InterPro; IPR002126; Cadherin.
 DR InterPro; IPR00233; Cadherin_C term.
 DR Pfam; PF00028; cadherin; 5.
 DR PRINTS; PR00205; CADHERIN.
 DR SMART; SM00112; CA; 4.
 DR PROSITE; PS00232; CADHERIN_1; 3.
 DR PROSITE; PS0268; CADHERIN_2; 4.
 DR Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
 KW Signal; Phosphorylation; 3D-structure.
 FT SIGNAL 1 23
 FT PROPEP 24 156
 FT CHAIN 157 884
 FT DOMAIN 157 709
 FT TRANSMEM 710 733
 FT DOMAIN 734 884
 FT DOMAIN 157 264
 FT DOMAIN 265 377
 FT DOMAIN 378 488
 FT DOMAIN 489 595
 FT DOMAIN 596 699
 FT DOMAIN 840 855
 FT MOD_RES 840 842
 FT MOD_RES 848 848
 FT CARBOHYD 560 560
 FT CARBOHYD 639 639
 FT CONFLICT 267 267
 FT CONFLICT 272 272
 FT STRAND 163 166
 FT TURN 167 168
 FT STRAND 175 179
 FT HELIX 183 186
 FT TURN 187 187
 FT STRAND 190 195
 FT STRAND 197 197
 FT TURN 198 200
 FT STRAND 201 201
 FT TURN 204 205
 FT STRAND 207 209
 FT TURN 211 213
 FT STRAND 215 218
 FT TURN 224 226
 FT STRAND 229 238
 FT TURN 239 240
 FT STRAND 243 243
 FT STRAND 248 255
 FT STRAND 263 264
 FT STRAND 268 274
 FT TURN 275 276
 FT TURN 279 280
 FT STRAND 282 285
 FT STRAND 288 289
 FT TURN 294 296
 FT TURN 299 300
 FT STRAND 303 310
 FT STRAND 319 321
 FT TURN 323 325
 FT STRAND 327 330
 FT TURN 337 339
 FT STRAND 342 350
 FT TURN 351 351


```

FT CONFLICT 3 3 S -> G (IN REF. 2).
FT CONFLICT 112 112 I -> N (IN REF. 2).
FT CONFLICT 573 573 S -> P (IN REF. 2).
FT CONFLICT 864 864 S -> P (IN REF. 2).
SQ SEQUENCE 880 AA; 97650 MW; C6CCDD91566427D86 CRC64;

Query Match
Best Local Similarity 54.4%; Score 2375; DB 1; Length 880;
Matches 471; Conservative 128; Mismatches 214; Indels 56; Gaps 9;

QY 11 LLLQVWMLQCAASEPCRAVFEAEVLEAGGAQEPQALGV-FMGCPGQBPALFSTD 69
Db 17 LCLLVVPSINADVSGCKGPFSSAEYFVSNRELEGRKLGVKNFSDCTTRKHGLYDVG 76
QY 70 NDFTV-----RNGETVQERS 86
Db 77 DSRPRVLPDGTVLKRVHKLHDKTKFTISTWDARGIKHSTNIASVAKRHSRSGEASRS 136
QY 87 LKERNPLKIPSKRI-IRRHKRWVAPISPENGKPPQRLNQLKSNKORDTKIPYSI 145
Db 137 ---KLPVLTPTTGTGLRKKRQWVPIPIKVSNERGPPFRLVQIKSNKDRFNKYYSI 193
QY 146 TPGADSPPGVFAVEKETGMLLNKLPDREIEAKYELFCHAVSENGASVEDPMNISIV 205
Db 194 TGGADNPPQGVFRIEWTGMLVTRDLREEDYKVLSSHAVSENGSPVEEPMETINV 253
QY 206 TDQNDHKPKTQDTFRGSVLGVLPGTSVMQVATDEDDAIYTYNGVWYSHSQPKDP 265
Db 254 IDQNDNRPKTQDFRGSVREGVQPGQVMAVSATDEDDNIDSLNGVLSYILKQDPPEP 313
QY 266 HDLFTTHRSSTISVSSGLDREKPEYTLTQATDMGDGTTTAVAVETILDANDNA 325
Db 314 IPNLTFTNRETGVLSLGTDLREKFEYTLTQATDLEGAGLSVSGKALIIQITDANDNA 373
QY 326 PMFDQKYEAEHPNAVGEVQRVLTVDLAPNSPAWRATYLMGGDDGHDFTTHPS 385
Db 374 PIFDKTYTALVPENEGFEVQRLSVTDLDMPGTPAWQVYKIR-VNEGFFNITTDPS 432
QY 386 NOGLTTRKGLDFAKQHTLYEVTEAEFVLKLPSTATIVVHVEDVNEAEVFPVPPSK 445
Db 433 NOGLTTRKGLDFAKQHTLYEVTEAEFVLKLPSTATIVVHVEDVNEAEVFPVPAVS 492
QY 446 VVEQEGITGPEVCVYTAEDPKEN-QKISYRILRDPAGLWAMDPSGOVTAVGTLDR 504
Db 493 RVDVSEDLRGEKISLVAQDPKQIQKLSYFIGNDPASWLTWNKDNQVITVGNGLDRE 552
QY 505 DEQVRNRIEVMVLANDNGSPPTGTGTLTLLIDVNDHGPVPEPROITICNSPVHV 564
Db 553 SE-YVKNTYVIMLVTDGVSFGTGTILHLVDVNDNGPVPSPRVFTMCDQNPQPV 611
QY 565 LNIIDKLSPTSFOALTDSDIYVTAENBEGDTVLVLSKKFLKQDTYDVHLSLSH 624
Db 612 LTISDAIPENTYKYVLSHSGDLTWKASLDSKGSMLSPQQLKKGDSYIVLLSDA 671
QY 625 GNEQLTVIRATVCDCHGVETCPGWKGFILP----VLGAVALLFLLVLLLVK 680
Db 672 QNNPQLTVNATVCSGCEKAKCOEKLVGGFDPILFLLVLSVALLFLLVLLFLK 731
QY 681 RKIPEILLPDDTRDNVYVYEGGEEDQDYDITOLHGLEARPEVLRNDVAPITP 740
Db 732 KVKPEILLPDDTRDNIFVYEGGEEDQDYDLSQHLGLSRDP-IMENDVVTLM 790
QY 741 TMYRPRANPDEIGNFIENLKAANTDPAPPYDTLLVDFYSGSDAASLSLSSAS 800
Db 791 APHYRPRSNPDEIGNFIENLKAANTDPAPPYDLSLLVDFYSGSEASLSLSSNS 850
QY 801 DQDQDYLNWGSGRFKKLADMYGGGDD 829
Db 851 NDEHDYVLSWGSFRKRLADMYGGDDDE 879

```

```

ID CADB_XENLA STANDARD; PRT; 884 AA.
AC P33152;
DT 01-OCT-1993 (Rel. 27, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Blastomere-cadherin precursor (B-cadherin) (XBCad).
OS Xenopus laevis (African Clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=83355;
RN [1]
RP SEQUENCE FROM N.A. PubMed=7531482;
RX MEDLINE=95151580; PubMed=7531482;
RA Mueller H.-A.J., Kuehl M., Finemann S., Schneider S.,
RA van der Poel S.Z., Hausen P., Wedlich D.;
RT "Xenopus cadherins: the maternal pool comprises distinguishable
RT members of the family.";
RL Mech. Dev. 47:213-223(1994).
RN [2]
RP SEQUENCE OF 438-884 FROM N.A.
RX MEDLINE=92062581; PubMed=1840622;
RA Herzberg F., Wildermuth V., Wedlich D.;
RT "Expression of XBCad, a novel cadherin, during oogenesis and early
RT development of Xenopus.";
RL Mech. Dev. 35:33-42(1991).
CC -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
CC They preferentially interact with themselves in a homophilic
CC manner in connecting cells; cadherins may thus contribute to the
CC sorting of heterogeneous cell types.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: Expressed in pituitary gland, lung and kidney.
CC -!- DEVELOPMENTAL STAGE: During oogenesis and early development.
CC -!- SIMILARITY: Contains 5 cadherin domains.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X78546; CAA55292.1; ALT_INIT.
CC EMBL; X63719; CAA45251.1; -.
CC HSP; P09803; 1SUH.
CC InterPro; IPR002126; Cadherin.
CC InterPro; IPR000233; Cadherin_C_term.
CC Pfam; PF00028; cadherin; 5.
CC Pfam; PF01049; Cadherin_C_term; 1.
CC PRINTS; P00205; CADHERIN.
CC SMART; SM00112; CA; 4.
CC PROSITE; PS00232; CADHERIN_1; 3.
CC PROSITE; PS00268; CADHERIN_2; 4.
CC Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
CC Signal.
KW SIGNAL.
FT PROPEP 1 26
FT CHAIN 27 157
FT DOMAIN 158 884 BLASTOMERE-CADHERIN.
FT TRANSMEM 158 706 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 707 730 POTENTIAL.
FT DOMAIN 731 884 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 158 265 CADHERIN 1.
FT DOMAIN 266 378 CADHERIN 2.
FT DOMAIN 379 489 CADHERIN 3.
FT DOMAIN 490 595 CADHERIN 4.
FT DOMAIN 596 706 CADHERIN 5.
FT DOMAIN 828 852 SER-RICH.
FT DOMAIN 879 884 ASP/GLU-RICH (ACIDIC).
FT CARBOHYD 427 427 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 560 560 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 683 683 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 438 440 ILT -> NSA (IN REF. 2).

```

FT	CONFLICT	677	677	R -> Q (IN REF. 2).
FT	CONFLICT	787	787	V -> A (IN REF. 2).
FT	CONFLICT	820	820	D -> N (IN REF. 2).
FT	CONFLICT	857	857	D -> N (IN REF. 2).
FT	CONFLICT	863	863	D -> N (IN REF. 2).
FT	CONFLICT	882	882	MISSING (IN REF. 2).
SQ	SEQUENCE	884 AA;	97980 MW;	98865D78E1DCB75B CRC64;
Query Match				
Best Local Similarity 54.3%; Score 2371; DB 1; Length 884;				
Matches 469; Conservative 125; Mismatches 222; Indels 51; Gaps 8;				
Qy	11	LLLLQVCMQCAASEPCRAVFAEAVTLEAGGAEQEPQALGV-FMGCPGQEPALFSTD	69	
Db	18	LCLLQVPSINVDVSGCQPGFSSANYTFSVNRLELRGRKLGKLVNVDCTTRKHGLYDVG	77	
Qy	70	NDDFTVRNGEIV-----QERRSLK	88	
Db	78	DSRFVLPDGTGLVKRHVLLKSHKDTREFTISTWDARGIKHSTNISVNVKHSRSGEARSR	137	
Qy	89	BRPLKIFPPSKRI-LRHKEDWVAPISVPENKGPFPQRLNOLKSKNDRDTKIFYSITG	147	
Db	138	SELPVLFPFKHGLKGLKKRWDWIPPIKVSNERGPPPKVLQIKSKNKLKSVFYSITG	197	
Qy	148	PGADSPPEGVFAVEKETGWLKLLNKLPLDREBIAKYELFGHAVSENGASVEDPMNISIVTD	207	
Db	198	QGADTPEGIFRIEKEKTHGWQVTEPLDREYEVKVLSSHAVSENGASVEEPMETVTVTD	257	
Qy	208	QNDHKPFTDTRGSLVGLVPGTSVMQVATDEDDAIYVNGVAVYSIHSEPKDPHD	267	
Db	258	QDNRPKFTQVPRGSRVREGVQGTKMSVSATDDSDSLNGVAYSLKODPEPIP	317	
Qy	268	LMFTIRSTGTISVISGLDREKVPEYTLTIQATDMGDGSTTTAVAVRILLDANNAFM	327	
Db	318	NLTINRETGVISLIGLDREKPEYTLVQAADLDGAGLTREGKAVIETDANNAFI	377	
Qy	328	FDQKYEAVHPENAVGHEVQRLTVDLDAPNSPAWRATYILIMGDDGDHFTITTHPESNQ	387	
Db	378	FDPKTYTALVPENEVGFVQRLSVTLDDMPGTAQAQVYKIR-VNEGPFENITTDPSNQ	436	
Qy	388	GILTRKGLDPEAKNQHTLVEVTNEAPFVLKLPSTATVHVHVEDVNEAPFVPSKV	447	
Db	437	GIULTAKGLDFKVKQVQITVENAVFSPVLPSTATVTVTVEDVNEAPFVPSKV	496	
Qy	448	EVOEGITGTPVCVYTAEDPKEN-QKISYRIILDPAGWLAMPDPSQVTVAGTGLDREDE	506	
Db	497	DVSEDLTRGEKIVSLVAQDPDKQIQKLSYFIGNDPARWLITKNGVITGNGLDRESE	556	
Qy	507	QFVRNNIYFVNLAMNGSPPTGTGTLTLTLIDVNDHGVPEPRQITICNOSPVRHVLN	566	
Db	557	-YVKNNTYVIMLVTDGVPVGTGTGLILHLVLDINDNGFVPSFRVFTMCDQNPQVLT	615	
Qy	567	ITDKLSPTSPPQAOLTDDSDIYVTAENVNEEGDVTVLSLKFLKQDVTVDVHLSDHGN	626	
Db	616	ITDADIPNTYFVSLSHSGSELTWAEILDSKTSNLSFTQQLKGDYSIYVLLADAQA	675	
Qy	627	KEQLTVIRATVCDCHGVETCPGPKWGGFILP-----VLGAVLALLFLLVLLLVKRRK	682	
Db	676	NRQLTVNATVCICEGKAICQEKLVAGFDPILVILGSLALLTLISLLLLFLKRRKV	735	
Qy	683	IKEPILLPEDTDNDVYFYEEGGSEEDDYDTOLHRLGLEAPPEVVLVNDVAPLIPTP	742	
Db	736	VKEPILLPEDTDNDIYFYEEGGSEEDDYDUSQLHRLGLDAPD-IMRNDVVPILMSVP	794	
Qy	743	MYRFRANPDEIGNFIENLKAANTDPTAPPYDTLLVFDYEGSGSAAISLISSTASQ	802	
Db	795	HYRFRPSNPDEIGNFIDENLDAANDPTAPPYDLSLVFDYEGSGSEAAISLINSNN	854	
Qy	803	DQDYDVLNWSGRFKLADMYGGEDD	829	
Db	855	EHDYNYLNDWGRFRKLADMYGGDDDD	881	

RESULT 9	CAD3_BOVIN	STANDARD;	PRT;	491 AA.
ID	CAD3_BOVIN			
AC	P19535;			
DT	01-FEB-1991 (Rel. 17, Created)			
DT	01-FEB-1991 (Rel. 17, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Cadherin-3 (Placental-cadherin) (P-cadherin) (Fragment).			
GN	CDH3 OR CDHP.			
OS	Bos taurus (Bovine).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Rutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OC	Bovidae; Bovinae; Bos.			
OX	NCBI_TaxID=9913;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=90360979; PubMed=2390969;			
RA	Liaw C.W., Cannon C., Power M.D., Kibonaka P.K., Rubin L.L.;			
RT	"Identification and cloning of two species of cadherins in bovine endothelial cells.";			
RL	EMBO J. 9:2701-2708(1990).			
CC	FUNCTION: Cadherins are calcium dependent cell adhesion proteins. They preferentially interact with themselves in a homophilic manner in connecting cells; cadherins may thus contribute to the sorting of heterogeneous cell types.			
CC	SUBCELLULAR LOCATION: Type I membrane protein.			
CC	SIMILARITY: Contains 5 cadherin domains.			

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				

DR	EMBL; X53614; CAA37676.1; -			
DR	PIR; S11694; IJBOCP.			
DR	InterPro; IPR002126; Cadherin.			
DR	InterPro; IPR000233; Cadherin_C_term.			
DR	Pfam; PF00028; cadherin; 3.			
DR	Pfam; PF01049; Cadherin_C_term; 1.			
DR	PRINTS; PR00205; CADHERIN.			
DR	SMART; SM00112; CA; 2.			
DR	PROSITE; PS00232; CADHERIN_1; 1.			
DR	PROSITE; PS00268; CADHERIN_2; 2.			
KW	Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.			
FT	NON_TER 1 1			
FT	DOMAIN <1 316	EXTRACELLULAR (POTENTIAL).		
FT	TRANSMEM 317 339	POTENTIAL.		
FT	DOMAIN 340 491	CYTOPLASMIC (POTENTIAL).		
FT	DOMAIN <1 102	CADHERIN 3.		
FT	DOMAIN 103 208	CADHERIN 4.		
FT	DOMAIN 209 314	CADHERIN 5.		
FT	DOMAIN 447 462	SER-RICH.		
FT	CARBOHYD 228 228	N-LINKED (GLCNAC. .) (POTENTIAL).		
SQ	SEQUENCE 491 AA; 54207 MW; 68EB0AE5918C4771 CRC64;			
Query Match				
Best Local Similarity 53.2%; Score 2323; DB 1; Length 491;				
Matches 434; Conservative 23; Mismatches 34; Indels 0; Gaps 0;				
Qy	339	ENAVGHEVQRLTVDLDAPNSPAWRATYILMGDDGDHFTITTHPESNQGLTKGLDF	398	
Db	1	ENTVSHEVQRLTVDLDAPNSPAWRATYIRIVGGDNGDHFTITTHPESNQGLTKGLDF	60	
Qy	399	EAKQHTLYVVTNEAPFVLKLPSTATVHVHVEDVNEAPFVPSKVVEQGIPTGEP	458	
Db	61	EAKQHTLYVEVINEVPEVVKLPSTATVTVVTVEDVNEPVPVPSKVIEQIGISTGEP	120	
Qy	459	VCVTAEDPDKENKISYRIILDPAGWLAMPDPSQVTVAGTGLDREDEQFVRNNIYEMV	518	
Db	121	ICAVTARPDKSGKISYRIILDPAGWLAMPDPSQVTVAGTGLDREDEQFVRNNIYEMV	180	

QY 519 LAMNGSPPTGTTGTTLLTLIDVNDHGPVPEPQITICNQSPVRHVLNITDKDLSPTSP 578
Db 181 LATDGSPTTGTGTTLLTMDINDHGPVPEPQITICNQSPVQVNLITDKDLSPTAP 240
QY 579 FOAQLTDSDIYTAENWEGDVTWLSKFLKQDYDVHLSLSDGNKEQLTVIRATVC 538
Db 241 FOAQLTHDSVYTAENYKGDVALSLKFLKQGEYDVHLSLSDGNKEQLTVIRATVC 300
QY 639 DCHGHVETCPGPKGGPILVGLAVLALLPLLLVLLVLRKKGKIKPEPLLPEDDTFRD 698
Db 301 DCHGNWVTCDPWTGFLPILGNALLVLLVFLVLRKKGKIKPEPLLPEDDTFRD 360
QY 699 FYGEGGEGEDQDYDTQLHRLGLEARPEVVLNDVAPTIIPTPMYRPRPANDPDEIGNFI 758
Db 361 FYGEGGEGEDQDYDTQLHRLGLEARPEVVLNDVAPSIIPTPMYRPRPANDPDEIGNFI 420
QY 759 IENLKAANTPTAPPYDLVDFYEGSGSDAASLSLTSSASDQDQDYDYNLWNGSRFKK 818
Db 421 IENLKAANTPTAPPYDLSLVDFYEGSGSDAASLSLTSSASDQDQDYDYNLWNGSRFKK 480
QY 819 LADYGGGGEDD 829
Db 481 LADYGGGGQDD 491

RESULT 10
CADI CHICK STANDARD; PRT; 887 AA.
AC P08641;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Epithelial-cadherin precursor (E-cadherin) (Cadherin-1) (Liver cell
DE adhesion molecule) (L-CAM).
GN CDH1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=89017248; PubMed=3174655;
RA Sorkin B.C., Hemperly J.J., Edelman G.M., Cunningham B.A.;
RT "Structure of the gene for the liver cell adhesion molecule, L-CAM.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:7617-7621(1988).
RN (2)
RP SEQUENCE OF 51-887 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=87204217; PubMed=3472238;
RA Gallin W.J., Sorkin B.C., Edelman G.M., Cunningham B.A.;
RT "Sequence analysis of a cDNA clone encoding the liver cell adhesion
RT molecule, L-CAM";
RL Proc. Natl. Acad. Sci. U.S.A. 84:2808-2812(1987).
CC -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
CC They preferentially interact with themselves in a homophilic
CC manner in connecting cells; cadherins may thus contribute to the
CC sorting of heterogeneous cell types. E-cadherin is a ligand for
CC integrin alpha-5/beta-7.
CC -!- SUBUNIT: Homodimer
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: Non-neural epithelial tissues.
CC -!- SIMILARITY: Contains 5 cadherin domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M22190; AAA82572.1; -.

DR EMBL; M22180; AAA82572.1; JOINED.
DR EMBL; M22181; AAA82572.1; JOINED.
DR EMBL; M22182; AAA82572.1; JOINED.
DR EMBL; M22186; AAA82572.1; JOINED.
DR EMBL; M22183; AAA82572.1; JOINED.
DR EMBL; M22194; AAA82572.1; JOINED.
DR EMBL; M22184; AAA82572.1; JOINED.
DR EMBL; M22185; AAA82572.1; JOINED.
DR EMBL; M22189; AAA82572.1; JOINED.
DR EMBL; M22193; AAA82572.1; JOINED.
DR EMBL; M22187; AAA82572.1; JOINED.
DR EMBL; M22192; AAA82572.1; JOINED.
DR EMBL; M22191; AAA82572.1; JOINED.
DR EMBL; M22195; AAA82572.1; JOINED.
DR EMBL; M16260; AAA82573.1; -.
DR PIR; A30201; IJCHCL.
DR HSP; P09803; ISUH.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR00233; Cadherin_C-term.
DR Pfam; PF00028; cadherin; 5.
DR Pfam; PF01049; Cadherin_C-term; 1.
DR PRINTS; P00205; CADHERIN.
DR SMART; SM00112; CA; 4.
DR PROSITE; PS00232; CADHERIN_1; 3.
DR PROSITE; PS0268; CADHERIN_2; 5.
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
KW Signal.
FT SIGNAL 1 26 POTENTIAL.
FT PROPEP 27 160
FT CHAIN 161 887 EPITHELIAL-CADHERIN.
FT DOMAIN 161 714 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 715 735 POTENTIAL.
FT DOMAIN 736 887 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 161 268 CADHERIN 1.
FT DOMAIN 269 381 CADHERIN 2.
FT DOMAIN 382 493 CADHERIN 3.
FT DOMAIN 494 599 CADHERIN 4.
FT DOMAIN 600 704 CADHERIN 5.
FT DOMAIN 844 855 SER-RICH.
FT CARBOHYD 291 291 N-LINKED (GLCNAC. . .).
FT CARBOHYD 346 346 N-LINKED (GLCNAC. . .).
FT CARBOHYD 564 564 N-LINKED (GLCNAC. . .).
FT CARBOHYD 643 643 N-LINKED (GLCNAC. . .).
FT CONFLICT 140 140 T -> M (IN REF. 1).
SQ SEQUENCE 887 AA; 97783 MW; F56ABA9779A94A40 CRC64;
Query Match 52.0%; Score 2274; DB 1; Length 887;
Best Local Similarity 52.7%; Pred. No. 7,2e-134;
Matches 461; Conservative 122; Mismatches 236; Indels 55; Gaps 13;
QY 8 LASLLLLQVWLOCAASEPCRAVFAEVTLEAGAEQEPQALGKV-FWGCPQGEPAIF 66
Db 16 LVLLLLQVCCRCDAAAPCGPFAAETFSVPQDSVAAGRELGRVSVFAACSGRPWAVY 75
QY 67 -STD-----NDD-----FTVRNGETVQERRS----- 86
Db 76 VPTDFRKVNGDVVSTKRLTYLGRKTSFTIYAQDAMGRHSARVTVGRHRRHHHHH 135
QY 87 -LKENPPLKI-PP--SKRLLRHRRKDVVVAISVPENKGFPPQRLNQLKKNKDRDTKIF 142
Db 136 HLQDTPPAVLTPPKHDPGFLRRQKEDWVIPPISCLNHRGYPYPMRLVQIKSNKDKESKY 195
QY 143 YSITGPGADSPEGVFAVEKKTGWLNLNKLPLDRBIAKYLEFGHAVSENGASVEDPMNIS 202
Db 196 YSITQGGADSPVPGFIETRETGLWLEVTQDLREKIDRYTLSSHAVSASGQVDPDPWEII 255
QY 203 IIVTDQNDHKPKFTQDTPTFRGSLVGLVPGTSMVQWATDEDDAIYTVNGVAVYHSQEP 262
Db 256 IIVTDQNDKPKVFIKEVFGVYIENAKPGTSVMTVNTATDADDVNTDNGIVSYIVSQQP 315
QY 263 KPHDLMTFIHRSTGTISVISGDLREKVPETLTQATMDGDSSTTTAVAVVEILDAN 322
Db 316 PRPHQPMFTIDPAKGIIISVLGGLDRETTPTNLTIVQATDQEGKLSNTAILEVTDA 375

SEQUENCE OF 149-169.

RP MEDLINE=91347911; PubMed=1879345;
RX Anguine B., Mueller A.H.J., Kellermann J., Hausen P.;
RT "Differential expression of two cadherins in *Xenopus laevis*.";
RL Development 111:829-844(1991).
CC -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
CC They preferentially interact with themselves in a homophilic
CC manner in connecting cells; cadherins may thus contribute to the
CC sorting of heterogeneous cell types. E-cadherin is a ligand for
CC integrin alpha-5/beta-7.
CC -!- SUBUNIT: Homodimer.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: Non-neural epithelial tissues.
CC -!- DEVELOPMENTAL STAGE: Appears in the embryonic ectoderm during
CC gastrulation when epidermal differentiation commences and it
CC disappears from the neural plate area upon neural induction.
CC -!- SIMILARITY: Contains 5 cadherin domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U04708; AAA93116.1; -;
DR EMBL; L29057; AAA61489.1; -;
DR EMBL; X75454; CAA53206.1; -;
DR HSPF; P09803; L50H.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR00233; Cadherin_C_term.
DR Pfam; PF00028; cadherin; 5.
DR Pfam; PF01049; Cadherin_C term; 1.
DR PRINTS; PR00205; CADHERIN.
DR SMART; SM00112; CA; 4.
DR PROSITE; PS00232; CADHERIN_1; 3.
DR PROSITE; PS0268; CADHERIN_2; 4.
DR Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
KW Signal.
FT SIGNAL 1 25 POTENTIAL.
FT PROPEP 26 148
FT CHAIN 149 872 EPITHELIAL-CADHERIN
FT DOMAIN 149 701 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 702 722 POTENTIAL.
FT DOMAIN 723 872 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 244 254 CADHERIN 1.
FT DOMAIN 358 368 CADHERIN 2.
FT DOMAIN 577 587 CADHERIN 3.
FT DOMAIN 710 721 POLY-LEU.
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 456 456 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 552 552 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 669 669 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 242 242 I -> V (IN REF. 3).
FT CONFLICT 332 332 N -> T (IN REF. 3).
FT CONFLICT 487 487 E -> R (IN REF. 3).
FT CONFLICT 501 502 AT -> CS (IN REF. 3).
FT CONFLICT 539 541 GNG -> EMA (IN REF. 3).
FT CONFLICT 548 548 K -> R (IN REF. 3).
FT CONFLICT 557 557 V -> G (IN REF. 3).
FT CONFLICT 567 567 P -> L (IN REF. 3).
FT CONFLICT 602 604 GFR -> EPQ (IN REF. 3).
FT CONFLICT 638 640 GQS -> DK (IN REF. 2).
FT CONFLICT 640 647 SILELRPK -> VYLSSDL (IN REF. 3).
FT CONFLICT 660 660 T -> A (IN REF. 3).
FT CONFLICT 770 770 A -> S (IN REF. 3).
FT CONFLICT 842 842 P -> S (IN REF. 2 AND 3).
FT CONFLICT 870 871 DE -> GED (IN REF. 3).
SQ SEQUENCE 872 AA; 96065 MW; 08637967EEFB4664 CRC64;
50.3%; Score 2198; DB 1; Length 872;

Query Match

323 DNAPMEDPKYEAHPENAVGHEVORLVTVTLDPAPNSPAWATATYLINGDDGDHFTITTH 382
376 DNIPFNPMYEGVVEENKPGTEVARLVITVDDAPGSPAWAVHYHKSGLDGAFTSID 435
383 PFSNOGILITRKGDFEAKNQHTLVVEVNEAPFVLKLPITSTATIVVHVEDVNEAPFVP 442
436 PSTNNGILKTAKGLDYETKSRVLDVVTYENKVPVLTSTASTASVLTVLDVNEPFPVP 495
443 PSKVVEVQEGITGPVPCVYTAEDPKB-NOKISYRIILRDPAGWAMDPSGQVAVGTIL 501
496 PIKRVGVPEDLPVGQOVTSYTAQDPDRDMOKITRYGSDPAGWLYTHPENGIVTATQPL 555
502 DREDEQFVRNNIYEVWVLAMDNQSPPTGTGTLTLLTIDVNDHGPVPEPQITICNOSP 561
556 DRESVHAI-NSYKAILAVDNGIPDTTGTGTLTLLQLQVNDNGPTPEPSFICSRQPE 614
562 RHVLNITDKLSPHTSPFOAQLTDDSDIYTAENVNEBGDTVVLSSKKFLKQDITYVHLSL 621
615 KOILSIVDKLPHTPYFPALEHSGSNMVTETRGQ-DELAAMLKKELEPGEYNI-FVKL 673
622 SHGKNEQLTVIRATVCDCHGVETCP--GPWKGF---ILPVLGAVLALLFLLVLLLL 675
674 TDSQKAQVTVKAVQVCEGTAKNCERSYIVGLGVPAILGLGLGILALLILLULLLL 733
676 LVKRRKIKPELLLPEDTRDNVFIYEGEGGEDDQDYITQLHRGLEARPEVVLNRDVA 735
734 FARRKKEKEPLLPEDDMRDNVNVDYEGEGGEDDQDYLSQLHRLGDARPEVI-RNDVA 792
736 PIIITPMYRRPANDPBGNIENLKAANTDTAPPYDTLLVFDYEGSGSDAASLSL 795
793 PPLMAAPQTRRPANDPBGNFIDENLKAADTDTAPPYDSLLVFDYEGGSEATSLSL 852
796 TSSASDQDQDYDLNENWSRFFKLADMYGGGEDD 829
853 NSSASDQDQDYDLNENWNRFFKLAEYGGGEDD 886

RESULT 11

CADI_XENLA STANDARD; PRT; 872 AA.

ID CADI_XENLA
AC P30944; Q91709;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Epithelial-cadherin precursor (E-cadherin) (Uvomorulin) (XTCAD-1).
OS *Xenopus laevis* (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; *Xenopus*.
OX NCBI_TaxID=8355;
RN (1)
RN MEDLINE=9524376; PubMed=76009366;
RP Levine E., Lee C.H., Kintner C., Gumbiner B.M.;
RT "Selective disruption of E-cadherin function in early *Xenopus* embryos
RT by a dominant negative mutant";
RL Development 120:901-909(1994).
RN (2)
RN SEQUENCE FROM N.A.
RN TISSUE=Tail bud;
RC MEDLINE=94368839; PubMed=8086449;
RX Tooi O., Fujii G., Tashiro K., Shiohara K.;
RA "Molecular cloning of cDNA for XTCAD-1, a novel *Xenopus* cadherin, and
RA its expression in adult tissues and embryos of *Xenopus laevis*.";
RL Biochim. Biophys. Acta 1219:121-128(1994).
RN (3)
RN SEQUENCE OF 149-872 FROM N.A.
RP MEDLINE=94363396; PubMed=8081882;
RA Broders F., Girault J.M., Simonneau L., Thierry J.P.;
RT "Sequence and distribution of *Xenopus laevis* E-cadherin transcripts.";
RL Cell Adhes. Commun. 1:265-277(1993).
RN (4)

Best Local Similarity 52.1%; Pred. No. 3.8e-129;
Matches 461; Conservative 125; Mismatches 229; Indels 70; Gaps 16;

QY 1 MGLPR-----GFLASLLLOVC-----WLOCASEPRAVFEAEVTLLEAGGABQEPGQAL 51
DB 1 MGLKRPWLLGAVVLLTIIQGGIAEWQ-----CRMGSFKYSLFLV-PKNLETDKAL 53
QY 52 GKV-FMCGPQEPALFDTNDFTVRNGETVQERRSLKERNPLKIP-----96
DB 54 GRVIFNCEGVRVQFASKDPNFEIHKDGTIVIKNPAKMDKNRKTFRVLAWETKGHVYET 113
QY 97 -----PSKRILRRKRDWVAPISVPENKGGPPPPQRLNOLK 132
DB 114 NITLKREGRHRQDLFGKSHHPKSTGLKQKQDWMVPPITVSENEKGPFPFRIVQIK 173
QY 133 SNKDRDTKIFYITGPGADPPGPGFAVEKXETGWLINLKPLDREETAKVELFCHAVSENG 192
DB 174 SVAKEVKVYSITGQADTPPGVFAIGREDGLANVTRDLREADINVLVFSHAYSSNG 233
QY 193 ASVEDPMNIIIVTDQNDHKPKFTQDTFRGSVLEGLVPGTSVMQVTTATDEDDAIYTYNGV 252
DB 234 ANVEDPMNIIIVTDQNDNDPVFTQSVFEGSVPEGSKPGTAVNTVSATDADSDVMYNGV 293
QY 253 VAYSIHSEPKDHLMTFTTHRSVTGTSIVSSGLDREKVPETLTITQATDMD-GDGSTTT 311
DB 294 IYSILNQEPKPEPTNKMTTHSHSGLISVLTGLDREKNPVYLTITQADGESGKDRITT 353
QY 312 AVAVEILDANMPDPQKYEAHVENAVGHEVQRLTVDLDAPNSPAWRATYILIMG 371
DB 354 ATALIIVMDINDNPVDFPTQYAKVPENEVGEVRLTVDDEDIETDANNAVYKIIG 413
QY 372 DGDHFTTHPESNOGILTRKGLDEAKNOHTLYVEVTNEAPFVLKLPSTSTATTIVHV 431
DB 414 NEANYFSIQT-DTGNIGLLTKVGLDVELKQYILSVITVINKANFVPLOTSTATTIVSV 472
QY 432 EDVNEAFVPPSVVQVQGIPTGEVPCVYTTAEDPDK-NQKISYRIILRDPAGLAMP 490
DB 473 EDVNEAFVPPSVVQVQGIPTGEVPCVYTTAEDPDK-NQKISYRIILRDPAGLAMP 532
QY 491 DSGQVAVGLDRDEGDFVRNVIYVWVLANMGSPPTTCTGCTLLTLLIDVNDHGVPP 550
DB 533 DNGIVTGNGLDRE-SKFLVNTYKVIILAADSGSPSATGTGLVNLNLLVDNDGPFLEP 591
QY 551 RQITICNQSPVRHVLNTDQLSPTSPFOAQLTDDSDIYVTAENVBEGTVV-LSLKXF 609
DB 592 QQESFCQKQDGFVFTIIDRLSNTYFYKALTGESNENWALVT--GQSILELRPKKE 649
QY 610 LKQDTPVHLSLGHKNEQLTVTRATVCHGHVETC--PGWPKGGF----ILPVILGAV 653
DB 650 LEICQYDVMITLDSFGLSNVTKLHIICQCDGPKMQCEKAAIAGLGLSATSIVGLGGI 709
QY 664 LALLFLALLVLLVKKRKEPILLPEDDTRDNVYVYEGEGGEDDQYDIQTLHGLE 723
DB 710 LALLFLALLVLLVKKRKEPILLPEDDTRDNVYVYEGEGGEDDQYDIQTLHGLE 769
QY 724 ARPEVLRLNDVAPTIPTMYRPPANPDEIGNFIENLKAANTDPTAPPYDILLVFDYE 783
DB 770 ARPDVI-RNDVAP-VLAAPQYRPPANPDEIGNFIENLKAANTDPTAPPYDILLVFDYE 827
QY 784 GSGDASLSLTSASDQDDYLNWGSRFKKLADMYGGGED 828
DB 828 GSGDASLSLTSASDQDDYLNWGSRFKKLADMYGGGED 872

RESULT 12

CAD2_HUMAN STANDARD; PRT; 906 AA.

AC P19022; Q14923;
DT 01-NOV-1990 (Rel. 16, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Neural-cadherin precursor (N-cadherin) (Cadherin-2).

GN CDH2 OR CDHN OR NCAD.

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. PubMed=2216790;
RX MEDLINE=91016946; PubMed=2216790;
RA Reid R.A., Hemperly J.J.;
RT "Human N-cadherin: nucleotide and deduced amino acid sequence.";
RL Nucleic Acids Res. 18:5896-5896(1990).
RN [2]
RP REVISIONS TO 341; 699 AND 705.
RA Reid R.A.;
RL Submitted (NOV-1990) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=92363956; PubMed=1500442;
RA Salomon D., Avalon O., Patel-King R., Hynes R.O., Geiger B.;
RT "Extrafunctional distribution of N-cadherin in cultured human endothelial cells.";
RL J. Cell Sci. 102:7-17(1992).
RN [4]
RP SEQUENCE OF 160-906 FROM N.A.
RX MEDLINE=90347462; PubMed=2384753;
RA Walsh F.S., Barton C.H., Putt W., Moore S.E., Kelsell D.,
Spurr N., Goodfellow P.N.;
RT "N-cadherin gene maps to human chromosome 18 and is not linked to the E-cadherin gene.";
RL J. Neurochem. 55:805-812(1990).
RN [5]
RP SEQUENCE OF 1-20 FROM N.A.
RX MEDLINE=95048366; PubMed=7959764;
RA Wallis J.A., Fox M., Walsh F.S.;
RT "Structure of the human N-cadherin gene: YAC analysis and fine chromosomal mapping to 18q11.2.";
RL Genomics 22:172-179(1994).
CC -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins. They preferentially interact with themselves in a homophilic manner in connecting cells; cadherins may thus contribute to the sorting of heterogeneous cell types. N-cadherin may be involved in neuronal recognition mechanism.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- SIMILARITY: Contains 5 cadherin domains.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X57548; CAA40773.1; -
EMBL; X54315; CAA38213.1; -
EMBL; S42303; AAB22854.1; -
EMBL; M34064; AAA03236.1; -
EMBL; Z27420; CAA81799.1; -
PIR; A38870; IJHUCN.
HSSP; P15116; INGI.
Genew; HGNC:1759; CDH2.
MIM; 114020; -
GO; GO:0007155; P:cell adhesion; TAS.
InterPro; IPR002126; Cadherin.
InterPro; IPR000233; Cadherin_C_term.
Pfam; PF00028; cadherin; 5.
Pfam; PF01049; Cadherin_C_term; 1.
PRINTS; PR00205; CADHERIN.
SMART; SM00112; CA; 5.
PROSITE; PS00232; CADHERIN_1; 3.
PROSITE; PS00268; CADHERIN_2; 5.
Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat; Signal.
KW SIGNAL.
FT SIGNAL. 1 23 POTENTIAL.

```
FT PROPEP 24 159
FT CHAIN 160 906
FT DOMAIN 160 724
FT TRANSMEM 725 746
FT DOMAIN 747 906
FT DOMAIN 160 267
FT DOMAIN 268 382
FT DOMAIN 383 497
FT DOMAIN 498 603
FT DOMAIN 604 714
FT DOMAIN 714 863
FT CARBOHYD 190 190
FT CARBOHYD 273 273
FT CARBOHYD 325 325
FT CARBOHYD 402 402
FT CARBOHYD 572 572
FT CARBOHYD 651 651
FT CARBOHYD 692 692
FT CONFLICT 12 12
FT CONFLICT 16 16
FT CONFLICT 196 196
FT CONFLICT 212 212
FT CONFLICT 357 357
FT CONFLICT 867 867
SQ SEQUENCE 906 AA; 99851 MW; 72DDC7B8B57C7AFC CRC64;

Query Match 39.2%; Score 1714; DB 1; Length 906;
Best Local Similarity 42.8%; Pred. No. 5.28-99;
Matches 391; Conservative 129; Mismatches 291; Indels 102; Gaps 27;

QY 6 GPLASLLQLVCLQCAASEP-----CRAVPREAEVLEAGAEQEPGQALGKV-FMGCP 59
DQ 6 GALRTLLPLLALLQASVEASGETALCKTGPE-DVYSAVLSKDVHGGQPLLVKFSNCN 64
QY 60 GQ-----BPALFSTNDND--FTVFN-----GETVOER----- 84
DQ 65 GKRKQVSESPADPKVDEGDMGVAVRSFPLSSSEHAKFLIYAQDKETQEKQWQVAVKLSL 124
QY 85 -----SLKERNPLK--IFP---SKRI--LRHRXRDVWVAPISVPENKGPFPQRLNQLK 132
DQ 125 PTLTEBSKSAEVEIVFPQFSKSHGQQRQKRWIPIPIINLPENSRGFPFQELVRIR 184
QY 133 SNKDRDKYFISITGPGADSPGEGVAFVKEKETGMLLANKPLDREELAKYSLFGHVSNG 192
DQ 185 SDRDKNLSRYSTGPGADQPGPTGIFTINPISGGLSVTKPLDREQIARPHRAVDING 244
QY 193 ASVEDPMNISITVDNDHKPKQDTFRGSVLEGLVPGTSVQVMTATDEDDAIYTVNGV 252
DQ 245 NQVENPIDIVINIDMNDNRPEFLHQVWNGTVPGSKPGTYVMTVTAIDADDP-NALNGM 303
QY 253 VAYSISQEPKPDHDLMTIHRSTGTISVSSGLDREKVPETLTITQATMDGDD---GST 309
DQ 304 LRYRIVSQAPSTPSPNNFTINNETGDIITVAAGLDREKVOQYTLITQATMEGNPTYGLS 363
QY 310 TTAVAVVEILDANDNAMPDPQKYEAVHPENAVGHEVQRVLTVDLDAPNSPAWRATYLM 369
DQ 364 NTATAVITVDVNDNPPETAMFTFYGEVPERNDVIVANLTVTKDQKQPHTPANNAVYIS 423
QY 370 GGDGDGHTITTHESNOGILITRKGLDPEAKQHTLYVEVNEAPEVLKLT---PTSTAT 426
DQ 424 GGDPTGFACTDPSNDGLVTVVKPIDFTENFVLTVAAENQVLAQIQHPPOSTAT 483
QY 427 IVHVEDNEAPVFPSPKVEVQEGIPTEGPEVCVYTAEDPDK-ENQKISRYLRDPAGW 485
DQ 484 VSVTVIDVNDNPFAPNPKIIRQEBGLHAGTMTLTFTAQDPDRYMQQNIYTKLSDPANW 543
QY 486 LAMDEDSGOYTAGTLDREDEQVRNNIYEVWMLAMDGSPPTGTGTLTLLTLDVNDHG 545
DQ 544 LKIDPVNGQITITAVLDRESN-VKNNIYNATFLASDNGIPPSGTTQIYLLDNDNA 602
QY 546 PVPEPQITICNOSPVRHVLNIT--DKDLSFHTSPFQAQUTDSDIY---WTAEVNEEGD 600
DQ 603 PQVLPQEAETC-BTEPDNSINITALDYIDPNAGPFAFDLPFSVTKRWNTI-TLNGD 660
```

```
QY 601 TVVLSLK-KFLKQDTYDVHLSLDHGN--KEOLTIVIRATVCDCHGH-----VETCPGPK 652
DQ 661 FAQLNLIKLEAGIYEVFIITDSNPKSNISIRVKVCQCDSDNGDCTDVRIVGAGL 720
QY 653 GGFILPVLGAVLALLFLLLVLLLV-----RKRKIKIPELLLPEDDTRDNVFFYG 702
DQ 721 G-----TGAIIAILLCIITLILVLMFVWMKRDKERQAKQLLIDPDDEVDNLIK 774
QY 703 BEGGGEEDODYDITQLHRLGLEAPPEV-----VLNDVAPTITPTMYRPPA--NPDEIG 755
DQ 775 BEGGGEEDQYDLSLQQLQDPTVEPDAIKPVGIRMDERP-IHABPOYFVRSAPHPGDIG 833
QY 756 NFIENLKAANTDPTAPPYDTLLVFDYEGSGSDAASLSLTSASDODODDYDYLNEWGR 815
DQ 834 DFINEGLKAAADNDPTAPPYDLSLLVFDYEGSGTAGSLSSLSNSSSGGEGDYDYLNDWGR 893
QY 816 FKXLDYMGGGED 828
DQ 894 FKXLDYMGGGDD 906

RESULT 13
CAD2_CHICK STANDARD; PRT; 912 AA.
AC P10288; Q90630;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-OCT-2001 (Rel. 40, Last annotation update)
DE Neural-cadherin precursor (N-cadherin) (Cadherin-2).
GN CDH2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88153917; PubMed=2831236;
RA Hattori K., Nose A., Nagatuchi A., Takeichi M.;
RT "Cloning and expression of cDNA encoding a neural calcium-dependent
cell adhesion molecule: its identity in the cadherin gene family.";
RL J. Cell Biol. 106:873-881(1988).
RN [2]
RP SEQUENCE OF 1-25 FROM N.A.
RC STRAIN=Cornish white rock Cockerel;
RX MEDLINE=97354288; PubMed=9210582;
RA Li B., Paradies N.E., Brackenbury R.W.;
RT "Isolation and characterization of the promoter region of the chicken
N-cadherin gene.";
RL Gene 191:7-13(1997).
CC -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
They preferentially interact with themselves in a homophilic
manner in connecting cells; cadherins may thus contribute to the
sorting of heterogeneous cell types. N-cadherin may be involved in
neuronal recognition mechanism.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- SIMILARITY: Contains 5 cadherin domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DQ EMBL; X07277; CAA30258.1; -.
DQ EMBL; U15563; AAB62980.1; -.
DQ PIR; A29964; IJCHCN.
DQ HSP; P15116; 1NCJ.
DQ InterPro; IPR002126; Cadherin.
DQ InterPro; IPR002333; Cadherin_C_term.
```


DR Pfam; PF00028; cadherin; 5.
 DR Pfam; PF01049; Cadherin C-term; 1.
 DR PRINTS; PRO0205; CADHERIN.
 DR SMART; SMO0112; CA; 5.
 DR PROSITE; PS00232; CADHERIN 1; 3.
 DR PROSITE; PS0268; CADHERIN 2; 5.
 DR Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
 KW Signal.
 FT SIGNAL 1 28
 FT PROPEP 29 164
 FT CHAIN 165 912
 FT DOMAIN 165 729
 FT TRANSSEM 730 752
 FT DOMAIN 753 912
 FT DOMAIN 165 272
 FT DOMAIN 273 387
 FT DOMAIN 388 502
 FT DOMAIN 503 609
 FT DOMAIN 610 720
 FT DOMAIN 869 884
 FT CARBOHYD 278 278
 FT CARBOHYD 330 330
 FT CARBOHYD 407 407
 FT CARBOHYD 578 578
 FT CARBOHYD 628 628
 FT CARBOHYD 657 657
 FT CONFLICT 21 21
 FT SEQUENCE 912 AA; 100464 MW; 98ASAC9DC1FFC489 CRC64;
 Query Match 39.0%; Score 1704.5; DB 1; Length 912;
 Best Local Similarity 42.6%; Pred. No. 2.1e-98;
 Matches 390; Conservative 132; Mismatches 284; Indels 109; Gaps 27;
 7 PLASILL-LQVCLQCASEP-CRAVPE-----AEVLEAGGAEQPGQALGV-FMGC 58
 14 PLALMLAALQAPKATCEDMLCKMGFPEDVHSAVSRSHVH-----GQPLNVRPQC 68
 59 PQQEPALF-STDNDDFTV-----RNGETVQERR----- 85
 69 DENRKIVGSEPEDFRVEDGVVVAERSFOLSAEPTFVVSARDKETQEWQMKVLTLP 128
 86 -----SLKRNPLK--IFPSKI-----LRRKRWVAVIPSPVNGKPPQRLNOL 131
 129 EPAFTGASEKQKXIEDIFFWQYKDSHLLKQKRDWVIPPINLPNSRGPPFQELVRI 188
 132 KSNKDRDTKIFYSITGPGADSPGPGVAVKETEGLMLLNLKPLDREBIKAYELFCHAVSEN 191
 189 RSDRDKSLSLSVYTGFGADQPTGIFINISGLSVTKFELDREQIASFHLRAHVDVN 248
 192 GASVEDPMNISIIYTDNDHKPKTQDTFRGSLVLEGVLPQTSVMQVATDEDDAIYTYNG 251
 249 GNQVENPDIIVINDMNDNRPEFLHQVYMWNGTVPEGKPGYVYVTVTAIDADDP-NAQNG 307
 252 VVAYSIIHQEPKOPHDLMTFTIRSTGTSVTSGLDREKVPXYTLTQATDMGD---GS 308
 308 MURYILSQSSPSFNNFTINNETGDIITVAAGLDREKVQYTLICATDMEGNPTVGL 367
 309 TTTAVAVVEILDANDNAPMFPQKYEAHVPAVNAVGVHVEVQRLTVDLDAPNSPAWRAVYLI 368
 368 SNTATAVITVDNDNRPPEFTAMTFYGEVPEVNRVDIVANLTVTDKQDPHTPAWNAVYOM 427
 369 MCGDDGDHFTITHPESNOGILITTKGLDDEAKQHTLYVEVTVNEAPPVLK---PTSTA 425
 428 TGGDPTGQFTIITDPSNDGLVTVVVKPDIIDFTNRMFVLTVAENQVPLAKGIOHPPOSTA 487
 426 TTVHVEDVNEAPVFPSPKVVVEQGIPTGPEVCVYTAEDPDKENQKIS--YKILRDKPA 483
 488 TVSIITVDNWSYFVFNPKLVQREQGLLAGSMLTFTTARDPDVRYMQOTSLRYSKLSDDPA 547
 484 GHLAMPDPSQVATVGTLDREDEQVRNIVYVWVLMNDGSPPTGTGTLTLLTLDVND 543
 548 NWLKDIPVNGQITTTAVLDRE-SIYVQNNMNNATFLASDNGIPSPGSGTGLQIYLLDND 606

QY 544 HGPVPEPQITICNOSPVRHVNLIT--DKOLSPHTSPFQALTDSDIY---WTAEVNEE 598
 DB 607 NAPQNPKEATTC-ETLQPNAINITAVDPDIDPAGPFAFELPDPSPSIKRWIT-VRIIS 664
 QY 599 GBTVVLSLK-KFLKQDVTYVHLSLSDHGN--KEQLTVIRATVCDCH-----GHVETCGP 650
 DB 665 GDHAQLSLRIRFLEAGIYDPIVITDSGNPHASSTSVLKVVCQCDINGDCTDVRIVGA 724
 QY 651 WKGGFILVGLAVLALLFLIALVLLLV-----RKKRKIKEPLLLPDDTRDNRVY 700
 DB 725 GIG-----TGATITAILLCIIILILVLMFVWKKRDKERQAKQLLIDPEDDVRDLK 778
 QY 701 YGEGGGEEDQDITITQHRGLEARPEV-----VLRNDVAPTIPTMYRPPA--NPDE 753
 DB 779 YDEGGGEEDQDYLSQLQPDTPVEPAIKPVGIRRLDERP-IHAEPQYVRSAAHPGD 837
 QY 754 IGNFIENLKANTDPTAPPYDTLLVFDYEGSGDAASLSLTSSASDQDQDYLDLNEWG 813
 DB 838 IGDFFNEGLKADNDPTAPPYDLSLLVFDYEGSGTAGLSLSLSSSSSGEGEQDYLDLNDWG 897
 QY 814 SRFKKLADMYGGGED 828
 DB 898 PRFKKLADMYGGGD 912
 RESULT 14
 CAD2_MOUSE
 ID_CAD2_MOUSE STANDARD; PRT; 906 AA.
 AC P15116; Q64260;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Neural-cadherin precursor (N-cadherin) (Cadherin-2).
 GN CDH2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBF_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89346748; PubMed=2762814;
 RA Miyatani S., Shimamura K., Hata M., Nagafuchi A., Nose A.,
 RA Matsunaga M., Hata K., Takeichi M.;
 RT "Neural cadherin: role in selective cell-cell adhesion";
 RL Science 245:631-635(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Tamura K.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
 RN [3]
 RP PARTIAL SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6;
 RX MEDLINE=92409532; PubMed=1528849;
 RA Miyatani S., Copeland N.G., Gilbert D.J., Jenkins N.A., Takeichi M.;
 RT "Genomic structure and chromosomal mapping of the mouse N-cadherin gene";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:8443-8447(1992).
 RN [4]
 RP DEVELOPMENTAL STAGE
 RC STRAIN=C57BL/6; TISSUE=Testis;
 RX MEDLINE=97033837; PubMed=8879495;
 RA Munro S.B., Blaschuk O.W.;
 RT "A comprehensive survey of the cadherins expressed in the testes of fetal, immature, and adult mice utilizing the polymerase chain reaction";
 RL Biol. Reprod. 55:822-827(1996).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 160-267.
 RX MEDLINE=95191680; PubMed=7885471;
 RA Shapiro L., Fannon A.M., Kwong P.D., Thompson A., Lehmann M.S.,
 RA Grubel G., Legrand J.-F., Als-Nielsen J., Colman D.R.,
 RA Hendrickson W.A.;
 RT "Structural basis of cell-cell adhesion by cadherins";

QY 704 EGGGEEDQYDITQHRGLEARPEV-----VLKNDVAPTIPTMYRPREA--NPDEIGN 756
 DB 776 EGGGEEDQYDLSQLQPDVPEPAIKPVGIRLDERP-IHAEFPQYVRSAAHPHGDIGD 834
 QY 757 FIENLKAANDPTAPYDILLVDFYGGSGDAAASLSLSSASDQDQDYDYLNEGSRF 816
 DB 835 FINEGLKAANDPTAPPYDLSLLVDFYGGSGTAGSLSSLSLSSSGGGQDQDYDYLNDGPRF 894
 QY 817 KKLADMTGGGDD 828
 DB 895 KKLADMTGGGDD 906
 RESULT 15
 CAD4_CHICK STANDARD; PRT; 913 AA.
 AC P24503
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cadherin-4 precursor (Retinal-cadherin) (R-cadherin) (R-CAD).
 GN CDH4.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 OC Gallus.
 OC NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Retina;
 RX MEDLINE=9129341; PubMed=1712604;
 RA Inuzuka H., Miyataki S., Takeichi M.;
 RT "R-cadherin: a novel Ca(2+)-dependent cell-cell adhesion molecule
 RT expressed in the retina.";
 RL Neuron 7:69-79 (1991).
 CC -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
 CC They preferentially interact with themselves in a homophilic
 CC manner in connecting cells; cadherins may thus contribute to the
 CC sorting of heterogeneous cell types. May play an important role in
 CC retinal development.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- TISSUE SPECIFICITY: Embryonic brain and neuronal retina.
 CC -!- DEVELOPMENTAL STAGE: Detected only after some degree of neuronal
 CC differentiation has taken place and persists at least up to the
 CC newly hatched stage.
 CC -!- SIMILARITY: Contains 5 cadherin domains.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D14459; BAA03356.1; .
 CC PIR; JH0424; IJCHCR.
 CC HSP; P15116; INCI.
 CC InterPro; IPR002126; Cadherin.
 CC InterPro; IPR000233; Cadherin_C_term.
 CC Pfam; PF00028; cadherin; 5.
 CC Pfam; PF01049; Cadherin_C_term; 1.
 CC PRINTS; PR00205; CADHERIN.
 CC SMART; SM00112; CA; 5.
 CC PROSITE; PS00232; CADHERIN 1; 3.
 CC PROSITE; PS00268; CADHERIN 2; 5.
 CC Cell adhesion; Glycoprotein; transmembrane; Calcium-binding; Repeat;
 KW Signal.
 KW SIGNAL 1 19 POTENTIAL.
 FT PROPEP 20 166 POTENTIAL.
 FT CHAIN 167 913 CADHERIN-4.
 FT DOMAIN 167 731 EXTRACELLULAR (POTENTIAL).
 FT

FT TRANSMEM 732 753
 FT DOMAIN 754 913
 FT DOMAIN 167 274
 FT DOMAIN 275 389
 FT DOMAIN 390 504
 FT DOMAIN 505 610
 FT DOMAIN 611 721
 FT DOMAIN 722 885
 FT CARBOHYD 280 409
 FT CARBOHYD 409 499
 FT CARBOHYD 554 554
 FT CARBOHYD 629 629
 FT CARBOHYD 658 658
 FT CARBOHYD 699 699
 FT VARIANT 652 652
 FT SEQUENCE 913 AA; 100885 MW; BD2BB9EC815DD6 CRC64;
 Query Match 38.8%; Score 1694; DB 1; Length 913;
 Best Local Similarity 41.5%; Pred. No. 9.3e-98;
 Matches 383; Conservative 135; Mismatches 286; Indels 118; Gaps 25;
 QY 11 LLLQVWLCQA-----SEPCRAVFEAEVLEAGAGAEQEGQALGV-EMGCPQGE 62
 DB 7 LLLVLLVWGSAAALNGDLTVRPTCKGFESEBDYTAFTVSNIME-GOKLLKVFNNCAGNK 65
 QY 63 PALFSTNDNDFTVR----- 76
 DB 66 GRYETNSILDFKVRADGTMYAVHQQVQMAKQILMVTAMDPTLGRWEAIVRFLVGEKLQ 125
 QY 77 -NGETVQERRS-----LKERNLKIPP-----SKILLRHKDWDVAPISVENGKGPQPQ 126
 DB 126 HNGHKPKGRKSGPVDLAQQSDTLPLMRHQSAKGLREKQKDWIPPIINVENRGPPQ 185
 QY 127 RLNLQKSKNRDRTKIFYITGPGADSPPEGVFAVEKETGWLILNKLPLDREIAYELFGH 186
 DB 186 QLVIRSKDKKEIHRYSITGVAGDQPMVEFSIDPVSGRMVYTFPMDREERASVHLRAH 245
 QY 187 AVSENGASVEDPMNISLIVTDNDHKKFTQTPRGSVLEGVLPQTSVQVATDEDDAI 246
 DB 246 AVDMGNKVENDILYIVIDMNDNRPFINQVNGSVDEGSKFTYVMTVANDADST 305
 QY 247 YVINGVAYSIHQSPKDPHDLMTFTHSTGTISVSSGLDREKVPEYTLTIQATMDGD 306
 DB 306 -TANGKVRIRIVTQTPQSPSQMFINSETGDIVTAAAGLDREKQQVQVWVQATDMEN 364
 QY 307 ---GSTTTAVAVVEILDANDNAPMFDPOKYEAHVPENAVGHVEQVRLTVTDLDAPNSPAWR 363
 DB 365 LNYGLSNTATAITITVDNDNPPEFTTSTYSGEVPEVNVVAVNLTVMDRQPHSPNN 424
 QY 364 AYILMGDDGDHFTITTHPSNQILITRKLDPKAKNQHTLVVEVNEAPFLKLP-- 421
 DB 425 AYIRISGDPGSHFTIRTDFTVNEGTVVKAVDYEMMRAPMLTVWVSNQAPLASCIGMS 484
 QY 422 -TSTATIVVHVDVNEAPVFPVPPSKVVEQGIPTGEPVCVVYTABDDPK-ENQKISYRIL 479
 DB 485 FQSTAGTIVSDVNEAPVFPVPPSKVVEQGIPTGEPVCVVYTABDDPK-ENQKISYRIL 544
 QY 480 RDPAGWLANDPSQGVATVGTLDREDEQFVRNIVYEMVWLANDNGSPPTGTGTLTLLI 539
 DB 545 SDPANWLNINATNGQITTAADVLDRES-YIKNVYEAFLAADNGIPPSAGGTGLQITLI 603
 QY 540 DVNDHGPVPEPQITICNGSPVRHVNIT--DKLSPTSPFQAQLTDDSDIY---WTAE 594
 DB 604 DINDNAPELLKPAQIC-EKPNLVININATAADADIDNPVFPVFPVPSVAVRKNWTI- 661
 QY 595 VNEEGTVVLSLK-KFLKQDQTVHLSLSDHONKE--QLTVIRATVDC--HGHVETCPG 649
 DB 662 TRLNGDYAQLSLRIMYLEAGVYDVPIIVTDSGNPPFYNTSIKVKVCPDENGDCDTTGA 721
 QY 650 PWKGGFILPVGLAVLALLFLVALLLV-----RKRRIKEPILLPDDDDTRNVF 699
 DB 722 VAAAGL---GTGATAILIICIIILLTWLLLVVWVKRREKREHTKQLLIDPEDDVRDNL 778

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 22:02:53 ; Search time 57 Seconds
(without alignments)

4588.852 Million cell updates/sec

Title: US-09-916-849A-1

Perfect score: 4369

Sequence: 1 MGLPRGPLASLLQLVCWLQ.....NEWSRFFKLADMYGGGDD 829

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25:**

- 1: sp_archaea:**
- 2: sp_bacteria:**
- 3: sp_fungi:**
- 4: sp_human:**
- 5: sp_invertebrate:**
- 6: sp_mammal:**
- 7: sp_muc:**
- 8: sp_organelle:**
- 9: sp_phase:**
- 10: sp_plant:**
- 11: sp_rodent:**
- 12: sp_virus:**
- 13: sp_vertebrate:**
- 14: sp_unclassified:**
- 15: sp_virus:**
- 16: sp_bacteriap:**
- 17: sp_archaeap:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3584.5	81.6	822	11	Q8BSL6
2	3560	81.5	821	11	Q8BRE1
3	3433.5	78.6	753	11	Q80VY6
4	2400.5	54.9	901	4	Q9UII7
5	2216.5	50.7	821	4	Q9UII8
6	1893	43.3	864	13	Q9QZ37
7	1702	39.0	906	4	Q8NI73
8	1701	38.9	906	11	Q8BSI9
9	1638	37.5	824	4	Q8NB64
10	1578.5	36.1	922	13	P79883
11	1308	29.9	711	13	Q7ZYV8
12	1272.5	29.1	729	13	Q7SZW2
13	1245	28.5	677	13	Q7SZW3
14	1221.5	28.0	690	13	Q7ZYV6
15	1127	25.8	566	4	Q9HIG6
16	1033	23.6	714	11	Q8VDK4

ALIGNMENTS

RESULT 1

Q8BSL6
ID Q8BSL6 PRELIMINARY; PRT; 822 AA.
AC Q8BSL6;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Cadherin 3.
GN CDH3.

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Forelimb;

RX MEDLINE=22354683; PubMed=12466851;

RA The FANTOM Consortium,

RA the RIKEN Genome Exploration Research Group Phase I & II Team;

RT "Analysis of the mouse transcriptome based on functional annotation of

RL Nature 420:563-573 (2002)."

DR EMBL; AK031265; BAC27327.1; -.

DR MGD; MGI:88356; Cdh3.

DR GO; GO:0005886; C:plasma membrane; IDA.

DR InterPro; IPR002126; Cadherin.

DR InterPro; IPR000233; Cadherin_C_term.

DR Pfam; PF00028; cadherin; 5.

DR PRINTS; PR00205; CADHERIN.

DR SMART; SM00112; CA; 4.

DR PROSITE; PS00232; CADHERIN_1; 3.

DR PROSITE; PS0268; CADHERIN_2; 4.

SQ SEQUENCE 822 AA; 90611 NW; 7C653D58210A595A CRC64;

Query Match

Best Local Similarity 81.6%; Score 3564.5; DB 11; Length 822;

Matches 681; Conservative 56; Mismatches 82; Indels 13; Gaps 5;

Q8Bgl1 mus musculu
Q8R490 rattus norv
Q8Bm92 mus musculu
Q86CZ9 homo sapien
Q8NS22 homo sapien
Q9dfsl xenopus lae
Q9dfso xenopus lae
Q8C7G6 mus musculu
Q77704 canis famil
Q8OWS7 mus musculu
Q8nl47 ciona intes
Q8ggh3 gallus gall
Q8vl68 mus musculu
Q7zyv7 gallus gall
Q93264 xenopus lae
Q86ud2 homo sapien
Q920m3 mus musculu
Q8C4K6 mus musculu
Q9hb01 homo sapien
Q9W6G5 brachydanio
Q9hb00 homo sapien
Q91838 xenopus lae
Q95YK9 ciona savig
Q8CB32 mus musculu
Q17281 botryllus s
Q86up1 homo sapien
Q8wnw5 sus scrofa
Q86up0 homo sapien
Q8ayd0 gallus gall

17 1031 23.6 714 11 Q8Bgl1
18 1021 23.4 714 11 Q8R490
19 1020.5 23.4 785 11 Q8Bm92
20 969.5 22.2 796 4 Q86CZ9
21 965.5 22.1 790 4 Q8NS22
22 953.5 21.8 792 13 Q9dfsl
23 953.5 21.8 792 13 Q9dfso
24 951.5 21.8 796 11 Q8C7G6
25 948.5 21.7 784 6 Q77704
26 944.5 21.6 788 11 Q8OWS7
27 941.5 21.5 798 5 Q9NL47
28 939 21.5 798 13 Q8QGH3
29 937.5 21.5 788 11 Q8Vl68
30 937 21.4 798 13 Q7zyv7
31 935 21.4 794 13 Q93264
32 926.5 21.2 794 4 Q86UD2
33 921 21.1 801 11 Q920M3
34 909 20.8 832 11 Q8C4K6
35 904.5 20.7 894 4 Q9HB01
36 903 20.7 490 13 Q9W6G5
37 898 20.6 840 4 Q9HB00
38 894 20.5 790 13 Q91838
39 874 20.0 910 5 Q95YK9
40 869.5 19.9 719 11 Q8CB32
41 837.5 19.2 906 5 Q17281
42 807.5 18.5 781 4 Q86UP1
43 805 18.4 782 6 Q8WNW5
44 792.5 18.1 819 4 Q86UP0
45 784.5 18.0 773 13 Q8AYD0

```

QY 1 MGLPRGFLA-SLLLOVCMLOCAASEPCRAVF-REAETVLEAGGAEQBPQALGKVFMC 58
Db 1 MELLSPHAFLLLLLOVCMLSVSEPYRAGFGEAGVTLEVEGTDLEPSQVLKVALAG 60
QY 59 PQEPPALFSTNDFTVRNGETVQERRSLKERNPLKIPSKRILRRHKRWVAPISVPE 118
Db 61 QGMEHA---DNGDIILNLTGRVQGGKAMHS-----PPTRILRRKRWVMPPIFVPE 110
QY 119 NGKGPFPORLNOLKSNKDRDTKIFYSITGPGADSPPEGVAEVEKETGMLLNKPLDRBEI 178
Db 111 NGKGPFPORLNOLKSNKDRGTKIFYSITGPGADSPPEGVTIEKESGMLLNHPLDREKI 170
QY 179 AKYELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKPTQDTPRGSVLEGVLPQTSVMQVT 238
Db 171 VKYELFGHAVSENGASVEBPNNISIIIVTDQNDKPKFTQDTPRGSVLEGVLPQTSVMQVT 230
QY 239 ATDEDDAIYVNGVAVYSIHSOEPKDPHDLMTIHRSTGTISVSSGLDREKVPYELTI 298
Db 231 ATDEDDAVNYNGVAVYSIHSOEPKDPHDLMTIHRSTGTISVSSGLDREKVPYELTV 290
QY 299 QATMDGSGSTTTAVAVVEILDANDNAPMDFPKYEAHVPEVNAVGHVQRLTVDLDPN 358
Db 291 QATMDGSGSTTTAAVAVVQILDANDNAPMDFPKYEAHVPEVNAVGHVQRLTVDLDPN 350
QY 359 SPAMRATYHILMGDDGDHFTITTHPESNOGILTRKGLDFAKNOHTLYVEVTNEAPFVL 418
Db 351 SPAMRATYHILVGDDGDHFTITTHPESNOGILTRKGLDFAKNOHTLYVEVTNEAPFV 410
QY 419 KLPTSTATIVHVEDVNEAPVFPVPSKVVEVQSGIPTGEPVCVYTAEDPKENOKISYRI 478
Db 411 KLPTATATVAVHVKOVNEAPVFPVPSKVVEVQSGIPTGEPVCVYTAEDPKEDOKISYTI 470
QY 479 LRDPAGLAMPDPSGQVAVGLTDRDQFVRNNIYEVVLMADNGSPPTTGTGLLLTL 538
Db 471 SRDPANLAVDPDSGQITAGILDRDQFVKNNYEVVLMATDSGNPTTGTGLLLTL 530
QY 539 IDVNDHGPVPERQITICNOSPVRHVLNITDKLSPTSPFOALTDSDIYWTAEVNEE 598
Db 531 TDINDHGPVPERQITICNOSPVPQVLNITDKLSNPSPPFOALTDSDIYWTAEVSEK 590
QY 599 GDTVLSLKFLKQDVTYDHLSDHGNKEQLTVIRATVCDCHGV-ETCPGPKGGFIL 657
Db 591 GDTVALSLKFLKQDVTYDHLSDHGNKEQLTVIRATVCDCHGVQVNDPCKPKGGFIL 650
QY 658 PVLGAVLALLFLLVLLVLRKRIKEPILLPEDDTRDNVYFYGGEGGEDDQDYDTQ 717
Db 651 PILGAVLALLFLLVLLVLRKRIKEPILLPEDDTRDNVYFYGGEGGEDDQDYDTQ 710
QY 718 LRHGLEARPEVLRNDVAPTIIPTPMYRPRANPDEIGNFIENLKAANTDPTAPPYDTL 777
Db 711 LRHGLEARPEVLRNDVAPTIIPTPMYRPRANPDEIGNFIENLKAANTDPTAPPYDSL 770
QY 778 LVFDYEGSGDAASLSSTASDQDQDYDYLNEWGRFCKLADMYGGGEDD 829
Db 771 LVFDYEGSGDAASLSSTASDQDQDYDYLNEWGRFCKLADMYGGGEDD 822

```

RESULT 2

```

Q8BRE1 PRELIMINARY; PRT; 821 AA.
AC Q8BRE1;
DT 01-WAR-2003 (TREMELrel. 23, Created)
DT 01-WAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Cadherin 3.
GN CDH3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;

```

```

RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium;
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AK045041; BAC32194.1; -.
DR MGI; MGI:88356; Cdh3.
DR GO; GO:0005886; C:plasma membrane; IDA.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR000233; Cadherin_C_term.
DR Pfam; PF00028; cadherin; 5.
DR PRINTS; PRO0205; CADHERIN.
DR SMART; SM00112; CA; 4.
DR PROSITE; PS00232; CADHERIN_1; 3.
DR PROSITE; PS00268; CADHERIN_2; 4.
DR SEQUENCE 821 AA; 90512 MW; 168356CSFB5CCED0 CRC64;

Query Match 81.5%; Score 3560; DB 11; Length 821;
Best Local Similarity 81.6%; Pred. No. 6.7e-241;
Matches 679; Conservative 56; Mismatches 83; Indels 14; Gaps 5;

QY 1 MGLPRGFLA-SLLLOVCMLOCAASEPCRAVF-REAETVLEAGGAEQBPQALGKVFMC 58
Db 1 MELLSPHAFLLLLLOVCMLSVSEPYRAGFGEAGVTLEVEGTDLEPSQVLKVALAG- 59
QY 59 PQEPPALFSTNDFTVRNGETVQERRSLKERNPLKIPSKRILRRHKRWVAPISVPE 118
Db 60 ----QGMHADNGDIILTRGTQGGKAMHS-----PPTRILRRKRWVMPPIFVPE 109
QY 119 NGKGPFPORLNOLKSNKDRDTKIFYSITGPGADSPPEGVAEVEKETGMLLNKPLDRBEI 178
Db 110 NGKGPFPORLNOLKSNKDRGTKIFYSITGPGADSPPEGVTIEKESGMLLNHPLDREKI 169
QY 179 AKYELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKPTQDTPRGSVLEGVLPQTSVMQVT 238
Db 170 VKYELFGHAVSENGASVEBPNNISIIIVTDQNDKPKFTQDTPRGSVLEGVLPQTSVMQVT 229
QY 239 ATDEDDAIYVNGVAVYSIHSOEPKDPHDLMTIHRSTGTISVSSGLDREKVPYELTI 298
Db 230 ATDEDDAVNYNGVAVYSIHSOEPKDPHDLMTIHRSTGTISVSSGLDREKVPYELTV 289
QY 299 QATMDGSGSTTTAVAVVEILDANDNAPMDFPKYEAHVPEVNAVGHVQRLTVDLDPN 358
Db 290 QATMDGSGSTTTAAVAVVQILDANDNAPMDFPKYEAHVPEVNAVGHVQRLTVDLDPN 349
QY 359 SPAMRATYHILMGDDGDHFTITTHPESNOGILTRKGLDFAKNOHTLYVEVTNEAPFVL 418
Db 350 SPAMRATYHILVGDDGDHFTITTHPESNOGILTRKGLDFAKNOHTLYVEVTNEAPFV 409
QY 419 KLPTSTATIVHVEDVNEAPVFPVPSKVVEVQSGIPTGEPVCVYTAEDPKENOKISYRI 478
Db 410 KLPTATATVAVHVKOVNEAPVFPVPSKVVEVQSGIPTGEPVCVYTAEDPKEDOKISYTI 469
QY 479 LRDPAGLAMPDPSGQVAVGLTDRDQFVRNNIYEVVLMADNGSPPTTGTGLLLTL 538
Db 470 SRDPANLAVDPDSGQITAGILDRDQFVKNNYEVVLMATDSGNPTTGTGLLLTL 529
QY 539 IDVNDHGPVPERQITICNOSPVRHVLNITDKLSPTSPFOALTDSDIYWTAEVNEE 598
Db 530 TDINDHGPVPERQITICNOSPVPQVLNITDKLSNPSPPFOALTDSDIYWTAEVSEK 589
QY 599 GDTVLSLKFLKQDVTYDHLSDHGNKEQLTVIRATVCDCHGV-ETCPGPKGGFIL 657
Db 590 GDTVALSLKFLKQDVTYDHLSDHGNKEQLTVIRATVCDCHGVQVNDPCKPKGGFIL 649
QY 658 PVLGAVLALLFLLVLLVLRKRIKEPILLPEDDTRDNVYFYGGEGGEDDQDYDTQ 717
Db 650 PILGAVLALLFLLVLLVLRKRIKEPILLPEDDTRDNVYFYGGEGGEDDQDYDTQ 709
QY 718 LRHGLEARPEVLRNDVAPTIIPTPMYRPRANPDEIGNFIENLKAANTDPTAPPYDTL 777

```


Db	710	LHRGLEARPVLRNDVVPTFTPTMYRPRPANPDICGNFIENLKAANTDPTAPPYDSL	769
Qy	778	LVFDYEGSGSAAISLSSSTSSASDQDDYDYLNGSGRFXKLADMYGGEDD	829
Db	770	LVFDYEGSGSAAISLSSSTSSASDQDDYDYLNGSGRFXKLADMYGGEDD	821
RESULT 3			
ID	Q80VY6	PRELIMINARY;	PRT; 753 AA.
AC	Q80VY6		
DT	01-JUN-2003	(TREMBLrel. 24, Created)	
DT	01-JUN-2003	(TREMBLrel. 24, Last sequence update)	
DT	01-OCT-2003	(TREMBLrel. 25, Last annotation update)	
DE	Hypothetical protein (Fragment).		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISUE=Embryo;		
RC	MEDLINE=22388257; PubMed=12477932;		
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,		
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,		
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,		
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,		
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,		
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,		
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,		
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,		
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,		
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,		
RA	Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,		
RA	Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,		
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,		
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,		
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,		
RA	Krzyszynski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,		
RA	Jones S.J., Marra M.A.;		
RT	"Generation and initial analysis of more than 15,000 full-length human		
RT	and mouse cDNA sequences."		
RL	Proc. Natl. Acad. Sci. U.S.A. 99:15899-16903 (2002).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	TISUE=Embryo;		
RA	Strausberg R.;		
RL	Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; BC052189; AAH52189.1; -		
DR	GO; GO:0016020; C:membrane; IEA.		
DR	GO; GO:0005509; F:calcium ion binding; IEA.		
DR	GO; GO:0007156; P:homophilic cell adhesion; IEA.		
DR	InterPro; IPR002126; Cadherin.		
DR	InterPro; IPR00233; Cadherin_C term.		
DR	Pfam; PF00028; cadherin; 5.		
DR	Pfam; PF01049; Cadherin_C term; 1.		
DR	PRINTS; PR00205; CADHERIN.		
DR	SMART; SM00112; CA; 4.		
DR	PROSITE; PS00232; CADHERIN_1; 3.		
DR	PROSITE; PS0268; CADHERIN_2; 4.		
KW	Hypothetical protein.		
FT	NON TER		
SQ	SEQUENCE 753 AA; 83300 MW; 414AFB4059CE94D2 CRC64;		
Query Match			
Best Local Similarity 78.6%; Score 3433.5; DB 11; Length 753;			
Matches 639; Conservative 51; Mismatches 43; Indels 1; Gaps 1;			
Qy	97	PKSEILRRKRDWVAPISVPENKGGFPFQRLNQLKSNKDRDKIFYSITGPGADSPPEG	156
Db	20	PPFILLRRKKEWVWPIFVPEKGGFPFQRLNQLKSNKDRDKIFYSITGPGADSPPEG	79
Qy	157	VFAVEKETGMLLNKPLDRBEIAKYELFGHAVSENGASVEDPMNISLIIVTDQNDHKPKFT	216

Db	80	VFTIEKESGMLLHMPDLDRKIVKELYGHAVSENGASVEEPWNISLIIVTDQNDHKPKFT	139
Qy	217	QDTFRGSVLEGLPGTSVMQVATATDEDDAIYTYNGVAVYSIHSQEPKDPHDLMTIHRST	276
Db	140	QDTFRGSVLEGLPGTSVMQVATATDEDDAVNTYNGVAVYSIHSQEPKDPHDLMTIHRST	199
Qy	277	GTISVSSGLDRKPKVEYTLITQATMDGSGSTTAVAVVELDANDNAPMDPKVEAH	336
Db	200	GTISVSSGLDRKPKVEYTLITQATMDGSGSTTAAVAVVELDANDNAPMDPKVEAH	259
Qy	337	VPENAVGEYQRLTVTDLDAPNSPAMRATYLMGGDDGDFHTITTHPESNQGLITTRKGL	396
Db	260	VPENAVGEYQRLTVTDLDVPNSPAMRATYLMGGDDGDFHTITTHPESNQGLITTRKGL	319
Qy	397	DPEAKNQHTLYVEVTNEAPFVKLPSTATIVVHVEDVNEAPVFPVPPSKVEVEGPIGT	456
Db	320	DPEAQDQHTLYVEVTNEAPFVKLPSTATIVVHVEDVNEAPVFPVPPSKVEVEGPIGT	379
Qy	457	EPVCVYTAEDPPDKENOKISVRLRDPAGWLAMPDPSGQVAVGTGLDRDEQCFVRNNIYEV	516
Db	380	ELVCIYTAQDPKEDQKISYTSRDPANLAVDPDSGQITAAAGILDREDSQFVKNVYEV	439
Qy	517	MYLMDNGSPPTGTGTTLLTLDVNDHGVPEPQIITCNQSPVHVNLNITDKLSPT	576
Db	440	MYLMDNGSPPTGTGTTLLTLDVNDHGVPEPQIITCNQSPVHVNLNITDKLSPT	499
Qy	577	SPPQALTDSDSYTYTAENBEGDVTVLKXFLKQDVTVDVHLSLSDHGNKQLTVIRAT	636
Db	500	SPPQALTDSDSYTYTAENBEGDVTVLKXFLKQDVTVDVHLSLSDHGNKQLTVIRAT	559
Qy	637	VDCCHGV-ETCPGPKWGGFIPLVGLAVLALLLVLVLLLVKRRKKEPILLPEDDTR	695
Db	560	VDCCHGV-ETCPGPKWGGFIPLVGLAVLALLLVLVLLLVKRRKKEPILLPEDDTR	619
Qy	696	DNVYVYGGEGGEGDQDYDITQHRGLARPVEVLNDVAPTIPTMYRPRPANPDEIG	755
Db	620	DNVYVYGGEGGEGDQDYDITQHRGLARPVEVLNDVAPTIPTMYRPRPANPDEIG	679
Qy	756	NFIENLKAANTDPTAPPYDILLVFDYEGSGDAASLSLTSSASDQDDYDYLNEWGR	815
Db	680	NFIENLKAANTDPTAPPYDILLVFDYEGSGDAASLSLTSSASDQDDYDYLNEWGR	739
Qy	816	FKKLADMYGGEDD 829	
Db	740	FKKLADMYGGEDD 753	
RESULT 4			
Q9ULI7			
ID	Q9ULI7	PRELIMINARY;	PRT; 901 AA.
AC	Q9ULI7		
DT	01-MAY-2000	(TREMBLrel. 13, Created)	
DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)	
DT	01-OCT-2003	(TREMBLrel. 25, Last annotation update)	
DE	E-cadherin.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Shibamoto S., Fukudome Y.;		
RL	"E-cadherin mutant."		
RL	Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.		
CC	-1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.		
CC	THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC		
CC	MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE		
CC	SORTING OF HETEROGENEOUS CELL TYPES (BY SIMILARITY).		
CC	-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).		
DR	EMBL; AB025106; BAA88957.1; -		
DR	HSSP; P09803; 1SUH.		
DR	GO; GO:0016020; C:membrane; IEA.		

Db 308 DPFLPKMFTNRTNTGTVISVVTGLDRESFTYTLVVQAADLQEGSLSTATAVITVD 367
QY 321 ANDNAPMEDPKQYEAHVAVENAVGVEQLRTVTDLDAPNSPAWRATYILMGDDGDHFTIT 380
Db 368 TNDNPFIFNP----- 377
QY 381 THPESNQILATRKGLDFAKNQHTLYVEVITNEAPFVLKLPSTSTATIVVHVEDVNEAPVF 440
Db 378 -----TT--GLDFAEQYILHVAIVNVVPEVSLTSTATVTDVLDVNEAPIF 425
QY 441 VPPSKVVEVQGIPTGEFVCVYTAEDPK-ENQKISYRILRDPAGWLAMPDSDGOVTAVG 499
Db 426 VPPEKRVESDFGQGBITSYTAQEPDTEMEQKITVIRMRDTANWLEINPDTGAISTRA 485
QY 500 TLDREDEQFVNNIYEVNVLAMDNGSPPTGTGILLTLIDVNDHGPVPEPRLITICNQS 559
Db 486 ELDRDPFHVKNSYTAIIATDNGSPVATGTGILLILSDVNNDAPPEPRTIPFCERN 545
QY 560 PVRHVLNTDKLSPHTSPOAQTTDDSDIYVTAEVNE-EGDTVVLKFKLPKQDYYVH 618
Db 546 PKPOVINIADLPNTSPFAELTHGASANWTIYNDPTQESILPKMALEVDGDKIN 605
QY 619 LSLSDHGNKEOLTIVRAVCDCHGVETC--PGMKGF-----ILPVLGVALALLFLLLV 672
Db 606 LKMDNQKQOVTTILEVSVCDCEGAAGVCRKAQVPEAGLQIPAILGILGGILLILL 665
QY 673 LLLVVRKKRIKEPLLPEDDTRDNVFFYEGEGGEDDQYDITQLHRLGLEARPEVVLRN 732
Db 666 LLLFLRRRAVKEPLLPEDDTRDNVFFYDEGGEGEDQDPLSQHLHGLDAREVT-RN 724
QY 733 DVAPTIPTMYRPRPANDPBGIFTEENKAAINTDTPAPPYDILLVDFEGSGSDAASL 792
Db 725 DVAPTLMSVPRYLPRPANDPBGIFTEENKAAINTDTPAPPYDILLVDFEGSGSEASL 784
QY 793 SSLTSSASDQDQDYDLNEMWSRFKCLADMYGGGEDD 829
Db 785 SSLNSESDDQDQDYDLNEMWNRKFLADMYGGGEDD 821
RESULT 6
Q90237
ID Q90237 PRELIMINARY; PRT; 864 AA.
AC Q90237;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE E-cadherin.
GN CDH1.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21270060; PubMed=11376490;
RA Babb S.G., Barnett J., Doedens A.L., Cobb N., Liu Q., Sorkin B.C.,
RA Yelick P.C., Raymond P.A., Marrs J.A.;
RT "Zebrafish E-cadherin: expression during early embryogenesis and
RT regulation during brain development.";
RL Dev. Dyn. 221:231-237(2001).
CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS: CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; AF364811; AAK52054.1; -.
DR ZFIN; ZDB-GENE-010606-1; cdh1.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0005194; F:cell adhesion molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0007156; P:homophilic cell adhesion; IEA.

DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR000233; Cadherin_C_term.
DR Pfam; PF00028; cadherin; 5.
DR Pfam; PF01049; Cadherin_C_term; 1.
DR PRINTS; PR00205; CADHERIN.
DR SMART; SM00112; CA; 4.
DR PROSITE; PS00232; CADHERIN_1; 3.
DR PROSITE; PS02268; CADHERIN_2; 4.
KW Calcium; Calcium-binding; Cell adhesion; Glycoprotein.
SQ SEQUENCE 864 AA; 95254 MW; 5ACA19538396397C CRC64;
Query Match 43.3%; Score 1893; DB 13; Length 864;
Best Local Similarity 47.0%; Pred. No. 6.5e-124;
Matches 394; Conservative 121; Mismatches 284; Indels 40; Gaps 9;
QY 27 CRAVFEAEVTLBAGABOEPQALQKV-FMGCPQEPALFSTDNDDFVTRNGETVQER 85
Db 29 CTFGELEEFVFKVHRHLHSGKRLGVTFNNDGRTLFLQSIDKRFEINDGTVTLLK 88
QY 86 SLK-----ERPLKIFPSKRI-----LRHKRDVVVAPIS 115
Db 89 QVTLHGHKVFVSHVANDSSGMKHTASVRVERVPAQVSSDDVLTKNKVKRGWIIPLIS 148
QY 116 VPENGKGFPPRLNQLKSNKDRDTKIFYISITGPADSPPEGVFAVEKETGWLILNKLPLR 175
Db 149 VSENSKGFPPMRLVQIKSDYAIETRLAYKITGEGADLDEKGIPTIDRLSGWVSVTOLLK 208
QY 176 EETAKYELFGHAVSENGASVEDPMNISITVTDNDHKPKFTQDTFSGVLEGLVPGTSYM 235
Db 209 EKASAKYLAHANGVADVTEKPMDIITVTDNDKPKVPTQNPNGNVPEALEKEVFM 268
QY 236 QVTATDEDDAIYTYNGVVAISHSQEPKDPHLMFTIHRSTGTISVISSGLDREKVEPT 295
Db 269 TVTATDADDKENTDADISVLIISQDPPSPKPMFAINPVSGISVLEKGLDREQWFRV 328
QY 296 LTIQATDMGDGSTTAVAVVEILDANDNAPDFDKYEAHVAVENAVGVEQLRTVTDLD 355
Db 329 LVTTATDMNGEGLSTGTAVITVTDSDNDNAPLFQSSYTSVPEHQVGEVAKLPVTDGD 388
QY 356 APNSPAWRATYILMGDDGDHFTITTHPESNGIILTRKGLDFAKNQHTLYVEVITNEAP 415
Db 389 EPESTAWSTKYQIAGDKGFFNISTGFSRLSGIITVTKPLDYETKQVILSVIVNDK 448
QY 416 FVLKLPSTATIVVHVEDVNEAPVFPVPSKVVVEGEGPTGEPVCVYTAEDPKKE-NQKI 474
Db 449 FVGPLPTSTATVTVNVKVDNPEPEPIKFKISRPNENLPVGSNLPFPFAIDPDTKKQNI 508
QY 475 SVRILRDPAGWLAMPDSDGOVTAVGTLDRDEDSQFVRNNIYEVNVLAMDNG-SPPTGTGT 533
Db 509 TVRIGNDPSDWLNI-TGSCQIQVKALDRESN-VADGKYKALILALDNDVESPATGTGT 566
QY 534 LLLTLIDVNDHGPVPEPRLITICNQSPVRHVLNITDKLSPHTSPPQALTDSDIYVTA 593
Db 567 LVIELQDVNDNAPVINERTIKLCNRESAPVLLSITDKLPPFAGPFKVEPQDTSKNMSV 626
QY 594 EVNEEGDVTVLSLKKFLKQDITDVHLSLDHGNKQLTVIRATVCDCHGHVETCPGPKWG 653
Db 627 FNETGHFLNIKPKSQLEGGYKVVLRVADREGESQENIIQASVCDCKGEAFQCTDKQA 686
QY 654 GF-----ILPVLGVALALLFLLLVLRKKRIKEPLLPEDDTRDNVFFYEGEGGEE 709
Db 687 GIPLEGVLGVGILLILLALLLMLFRRKNSKEPLLPEDDVDRDNIYYDESGGED 746
QY 710 DDYDITQLHRLGLEARPEVVLNDVAPTIPTMYRPRPANDPDEIGNFIENLKAANTDP 769
Db 747 DQDFDLVHLRGLDNRPE-VFRNDVAPTFMPAPQYRPRPANPEIETGTTIDNKLKADNDP 805
QY 770 TAPPYDILLVDFEGSGSDAASLSTSSASDQDQDYDLNEMWSRFKFLADMYGGGEDD 828
Db 806 TAPPYDILLVDFEGSGSDAGSLSSINTSSGNDQDYDLNEMWGRFVKFLADMYGGGEDD 864
RESULT 7

Query Match 38.9%; Score 1701; DB 11; Length 906;
Best Local Similarity 42.6%; Pred. No. 2.1e-110;
Matches 391; Conservative 126; Mismatches 293; Indels 108; Gaps 28

```
QY 2 GLPRGLASILLLOVCWNLQCAASEP-----CRAVFEAEVTLRAGGAQEPGQALGKV-F 55
Db 6 GAPR-----TLPLLAALLQASVETSGEIALCKTGFFE-DVYSAVLPKDVHGQPLLNKVF 60
QY 56 MGC-----PGQBPALFSTDND--FTVRN-----GETVQER----- 84
Db 61 SNCNRKRVQYSESEPADFKYDEGTGVAVRSPFLTAQAKFLIYAQKETQEKQVAVN 120
QY 85 -----RSLKERNPLK--IPFSK-----RI LRHKRDWVAVIPSPVNGKGPFFQRL 128
Db 121 LSREFTLTPEEMKEFHEIEEVFPRLAKHSGALQKQKRDWVIPPINLPNSRGPFFQEL 180
QY 129 NOLKSNKDRDKIFVSTGPGADSPPEGVFAVEKETGWLILNKPLDREETAKYELFGHAV 188
Db 181 VRISDRXNLISLRYSVTGPGADPTGIFINISQGLSVTKEDRELLARFLRAHAV 240
QY 189 SENGASVEDPMNIIIVTDQNDHKPKFTQDFFRGSLVGLPGTSVNMQVATDEDDAIYT 248
Db 241 DINGNOVENPDIIVINVDMDNRPFLHQWNGSVPEGSKPGTYVMTVTAIDADDP-NA 299
QY 249 YNGVVAYSIIHQEPKDHLDLMTHTSRGTSTISVSSGLDREKYPEYTLITQATMDGD-- 306
Db 300 LNMRLVRILSQASTSPNNFTINNETGDIITVATGLDREKVOQYTLITQATMDGNT 359
QY 307 -GSTTTAVAVVEILDANDNAPMDFPKYEAHVPENAVGVHEVQRLTVTDLDAPNSPAWRAT 365
Db 360 YGLSNTATAVITVDNDNRPETAMTFYGEVPENRVDVIVANLTVTDKQPHTPANAA 419
QY 366 YLINGGDGDHFTTTHPESNOGLITRKGLDPRKQNTLYVEVNEAPFVKL-----PT 422
Db 420 YRISGGPTGFAITLDPNSDGLVTVVKPFDFTFMRMFLTVAAEQVFLAKGIQHPQ 479
QY 423 STATIVVHVEDNEAPVFPVPSKVVEQGIPTCEPVCVVTAEDPDK-ENQKISYRILRD 481
Db 480 STAIVSVTIVDVENPFAFNPKIIRQEBGLHAGTMTTLTQADPRYMOQNTIRYKLS 539
QY 482 PAGWLAMPDSQGVAVTGLDREDEQVRNNIYEVWVLAMDNPSPTGTGTLTLTLIDV 541
Db 540 PANWLKIDPVNGQITTTAVLDRESNP-VKNNIYNATFLASDNGIPPMGSGTGLIYLLDI 598
QY 542 NDHGPVPEPQITICNOSPVRHVLNIT--DKDLSPTSPFQAQLTDDSDIY---WTAENV 596
Db 599 NDNAPQVLPQEAETC-ETPEPNSINITALDYDIDPNAGPFAFDLPSPVTKRNT--IN 655
QY 597 E-EGDTVVLISL-KFLQDQTYDVHLSLSDHGN--KEQLTVIRATVCDCHGH-----VETC 647
Db 656 RLNGDPAQLNLKIKFLEAGIVEVPLIITDSGNPPKSNISILRVKVCQDSNGDCTVDRI 715
QY 648 PGPKWGGFIPVLGAVLALLPLLLVLLLY-----RKRKIKPELILLPEDDTRDN 697
Db 716 VGAGLG-----TGAIITAILLCIILLILVLMFYVMKRRDKEROAKQLLIDPEDDVRN 769
QY 698 VFYVYEGEGGEEDQDYDITOLHRLGLEARPEV-----VLRNDVAPTIIPTMYRPERA--N 750
Db 770 ILKYDESGGGEEDQDYLSQLQPDVTEPDALXPVGIRLDERP-THAEQYVPRGAAPH 828
QY 751 PDEIGNFIENLKAANTDPAPPYDTLLVFDYEGSGDAAASLSLSSASDQDQDYDLN 810
Db 829 PGDIGDFINEGLKAANDPTAPPYDLSLLVFDYEGSGTAGSLSSLSNSSSGGQDYDLN 888
QY 811 EWGSRFKKLADWYGGGBD 828
Db 889 DWGPRFKKLADWYGGGDD 906
QY 889 DWGPRFKKLADWYGGGDD 906
PRELIMINARY; PRT; 824 AA.
RESULT 9
Q8NB64
ID Q8NB64 PRELIMINARY; PRT; 824 AA.
AC Q8NB64;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ34177.
```

```
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,
RA Houta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J.,
RA Matanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
RA Sugiyama T., Irie R., Otsuki T., Sato H., Ota T., Wakamatsu A.,
RA Ishii S., Yamanoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; AK091496; BAC03677.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0005194; F:cell adhesion molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR000233; Cadherin_C_term.
DR Pfam; PF00028; cadherin; 5.
DR PRINTS; PR00205; Cadherin_C term; 1.
DR SMART; SM00112; CA; 5.
DR PROSITE; PS00322; CADHERIN_1; 3.
DR PROSITE; PS00268; CADHERIN_2; 3.
KW Hypothetical protein; Calcium; Calcium-binding; Cell adhesion;
KW Glycoprotein.
SQ SEQUENCE 824 AA; 90261 MW; 31A0C127BD40BAEB CRC64;
Query Match 37.5%; Score 1638; DB 4; Length 824;
Best Local Similarity 43.0%; Pred. No. 4,7e-106;
Matches 360; Conservative 135; Mismatches 271; Indels 72; Gaps 23;
QY 41 CGAEQEPGQALCKVFMGCPGQBPALFSTDNDFTVRNGETVQERSLKERNP-----L 93
Db 10 GGREN-----AGVTSKSCVGTGQYETSSP-----HSGHKPKQKGVVALDSPPKDTL 60
QY 94 KIFPSKRI---LRRHKRDWVAVIPSPVNGKGPFFQRLNQLKSNKDRDKIFYSITGPGA 150
Db 61 LPWPQHONANGLRRKRKRDWVIPPINVPNSRGPFFQQLVTRSKDNDIPRYSITGVA 120
QY 151 DSPPSGVFAVEKETGWLILNKPLDREETAKYELFGHAVSENGASVEDPMNIIIVTDQND 210
Db 121 DQPFMEVFSIDMSGRMTVTRPMDREEHASVHLRAHAVDMNGNKVENFIDLIYIYVDND 180
QY 211 HKPKFTQDFFRGSLVGLPGTSVNMQVATDEDDAIYTVNGVAVYSIHSQBPKPDPHLMF 270
Db 181 NRPEINQVYNGSVDEGSKPGTYVMTVTAIDADDST--TANGMVRVRIYVITQPSQSNWF 239
QY 271 TIHRTGTGISVSSGLDREKYPEYTLITQATMDGD---GSTTTAVAVVEILDANDNAPM 327
Db 240 TINSETGDIIVTAAGLDREKVOQYTVIQAOTDMEGNLYGSLNTATAILITVDNDNPE 299
QY 328 FDPQKYEAHVPENAVGVHEVQRLTVTDLDAPNSPAWRATYILMGDDGDFHTTTHPESNQ 387
Db 300 FTASTFAGEVPENRVETVVANLTVMDRQPHSPNNNAVYRIISGDPSPGFSVRTDPVNE 359
QY 388 GLLTTRKGLDPEAKNQHTLYVEVNEAPFVKLP---TSTATTIVVHVEDNEAPVFPVPS 444
Db 360 GMVTVVKAVDYELNARFMLTVWVSNQAPLASGIQMSFQSTAGVTISIMDINEAPVPSNH 419
QY 445 KVVEVQEGIPTGPEVPCVVTAEDEPK-ENQKISYRILRDPAGWLAMPDSQGVAVTGLDR 503
```

```

Db 420 KLIRLEGGVPPGTVLTTFSAVDPRFQQAVRYSKLSDFASWLHINATNGQITTAVALDR 479
QY 504 EDEGFVNNIYEVNVLAMNGSPPTCTGTLTLLTLDVNDHGVPPBPQITTCNQSPVRH 563
Db 480 E-SLYTKNNVYEAFLAAGNIPASGTGLQIYLDINDNAPELLPKQAIC-ERNLN 537
QY 564 VLNIT--DKDLSPTSPPQALTDSDIYMTAEVNE-----EGDTVWLSLK-KFLKQD 613
Db 538 AINITAADADVDPNIGFVPELP-----FVPAARVKNTWITRLNGDYAQLSLRLYLEAG 592
QY 614 TYDVHLSLSDHGNK--EQLTVIRATVDC--HGHVETCPGPWKGGFTLPVGLAVLALLFL 669
Db 593 MYDPVPIIVTDSGNPPLNTSIIKVKVCPDONGDCTTIGAVAAAGL---GTGAIVALLIC 649
QY 670 LVLVLLV-----RKKRKIKEPLLPPEDDTRDNVFFYGBEGGGEEDQDYITQLH 719
Db 650 ILIILLTWLLFVWMMKREKERTKQLLDPEDDVRNLIKYEDEGGGGEEDQDYLSQLQ 709
QY 720 RGLRARPV-----VLNRVAPTIITPMY--RPRPANTDEIGNFIENLKAANTDPTA 771
Db 710 QP-EAMGHVPSKAPGVRRVDRP-VGAEPQYPIRPMVPHFPGDIDGDFINEGLRAANDPTA 767
QY 772 PPTLLAVDFYEGSGSDAASLSLTSSASDQDQDYVLYNEWGSFRFKLADMYGGGEDD 829
Db 766 PPDLSLLVDFYEGSGSTAGSVSSLSNSSSG--DQDYDLYNDWGPFRFKLADMYGGGEDD 824

```

RESULT 10

```

P79883
ID P79883 PRELIMINARY; PRT; 922 AA.
AC 01-MAY-1997 (Tremblrel. 03, Last sequence update)
DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Maternally expressed neural cadherin Xmn-cadherin.
GN MN-CADHERIN.
OS Xenopus.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae.
OC NCBI TaxID=8353;
RX MEDLINE=96257964; PubMed=8652409; Koga C., Ito Y., Hikasa H.,
RA Tashiro K., Tooi O., Nakamura H.,
RA Shikawa K.;
RT "Cloning and expression studies of cDNA for a novel Xenopus cadherin
RT (Xmn-cadherin), expressed maternally and later neural-specifically in
RT embryogenesis.";
RL Mech. Dev. 54:161-171(1996).
CC !- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES (BY SIMILARITY).
CC !- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; S82457; AAB37685.2; -.
DR HSSP; P15116; UNCT.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0005194; F:cell adhesion molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR000233; Cadherin_C_term.
DR Pfam; PF00028; cadherin; 5.
DR PRINTS; PF01049; Cadherin_C_term; 1.
DR SMART; SM00112; CA; 5.
DR PROSITE; PS00232; CADHERIN_1; 3.
DR PROSITE; PS0268; CADHERIN_2; 3.
KW Calcium; Calcium-binding; Cell adhesion; Glycoprotein.

```

```

SQ SEQUENCE 922 AA; 101173 MW; FEA7A95CBAF1B640 CRC64;
Query Match 36.1%; Score 1578.5; DB 13; Length 922;
Best Local Similarity 41.4%; Pred. No. 8.4e-102;
Matches 364; Conservative 127; Mismatches 275; Indels 113; Gaps 25;
QY 48 GOALGKV-FMCC-PQOEPAFLSTNDPFTVNGETVQERRSLK----- 88
Db 60 GQRLKLVFNGSSGAQGIWYETNNPDKVGADGAVYTAREVOIPAKQAKFIVAANDHET 119
QY 89 -----ERNPLK-----IPPSK-----RIURHKRWVV 111
Db 120 PEKWEAAIQLFVEETPLNQSOVQTGSQSESEQSQSGTLLPWRQHHKGLRRQKRDWVI 179
QY 112 APISYPENKGGFPQRLNQLKSNKDRDKIFYSITGCGADSPPEGVFAVEKETGMLLNK 171
Db 180 PFVNVPSRSGFPQOOLVLRSDKDRDRTIYSITGVGADQPPMAIFNIDIFGRMNVTR 239
QY 172 PLDRBEIAKYELFGHAVSENGASVEDPMNIIIVTDQNDHKPKFTQDTFRGSVLEGLPG 231
Db 240 PLDRERSYHLRAHVDINGNKVENPIDLSIYVIMDMNDRPEFSSPIFNGSVDEASKPG 299
QY 232 TSMQVATDEDDALITYNGVVAISHQBPQKPDHLMFTTHRSTGTISVSSGLDREKV 291
Db 300 TYVMTVTAHDADD-INTSGIVMYRIMDQSPQSPSHDMFVHSKTGVTINTVAAGLDREKV 358
QY 292 PEYTLTIQATDMGDP---GSTTTAVVVEILDANDNAPMFPQKYEAHVPENAVGEYQR 348
Db 359 QQTIVVIQATMEGNLHGLSNTATITVADVNDNPEFTKRMFIGEVPENHVDVVVAN 418
QY 349 LTVTLDAPNSPAWEATYILMGDDGDHFTTTHPESNOGILTRKGLDFPAKQHTLYV 408
Db 419 LTVVDRDQPYTSNMNAVFKIISGDPDGHFTIKTDEVTNEGIVTSKPYDYMVKVPLIV 478
QY 409 EYTNAPFVKLP---TSTATVHVHEDVNEAPVFPVPSKVVEQEGITGEPVCVYTA 465
Db 479 MYTNOAPLASIQMSLOSTAAVTVSVNDVNEAPYFPNREPIRKLEGSAGRLITTSV 538
QY 466 DPKENQKI-SYRIILRDPAGLAMPDPSQVTVAGTGLDREDEQFVRNNIYEVNVLAMNG 524
Db 539 DPHSNQQVLYRYSKISDPANLAINTTNGQVSTTAVLDRE-SPFVKDLYQAKFLATDNG 597
QY 525 SPPTGTGTLTLLTLDVNDHGVPPBPQITTCNQSPVRHVNIT--DKDLSPTSPPQAA 582
Db 598 NPPASGTGLLIQI-LINDNAPELLPKQAIC-ERNPENGINITAIDVDRPSADPFVFE 656
QY 583 LTDDSDIY-----WTA-EVNERGDTVWLSLKFLKQDQDYDVHLSLSDHGNKE--QLTVIR 634
Db 657 L--PSVPYTIERNWTIHRINSIYARLSLIQI-GYLESGMYDVFIIVTDSGNPLVNTSIK 713
QY 635 ATVDC--HGHVETCPGPWKGGFTLPVGLAVLALLFLVLLVLLV-----RKKRK 682
Db 714 VKVPCDNGDCTTIGAVAAAGL---GTGAISILICIIILLSMVLLFVWMMKREKERH 770
QY 683 IKEPLLPEDDTRDNVFFYGBEGGGEEDQDYITQLHRLGLEARPEV-----VLNR 732
Db 771 TQQLLDPEDDVRNLIKYEDEGGGGEEDQDYLSQLQ-----QPETLDHVLNKKVAGVRV 825
QY 733 DVAPTIPTMYRPRPA--NPDEIGNFIENLKAANTDPTAPPYDITLLVDFYEGSGSDA 790
Db 826 DERP-VGAHQYPIRPMVPHPGDIDGDFITEGLRAANDNPTAPPYDLSLLVDFYEGSGSTAG 884
QY 791 SLSLTSSASDQDQDYVLYNEWGSFRFKLADMYGGGEDD 829
Db 885 SVSSLSNSSS-EDQDFDLYNDWGPFRFKLADMYGGGAED 922

```

RESULT 11

```

Q7ZY98
ID Q7ZY98 PRELIMINARY; PRT; 711 AA.
AC Q7ZY98;
DT 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)

```



```

533 TLLATLIDVNDHGPVPEPRQITICNQSPVRHVNLITDKLSPTSFPQAQLTDDSDIYWT 592
540 TLVTELOQVNDNAPVNERIKLCNRESAPVLSITDNDGPPAGPFTVBPQDTSKWS 599
593 AENVNEEGDVTYLSLKFLKQDVTYVHLSDHGNKEQLTVIRATVCDCHGHVETCPGPK 652
600 VFNETHYLSLKPSQOLEQREYNVLRVTDREGQSQESAIOASVCDCKGEAPQCTNKQV 659
653 GGF----ILPVLGAVALLFLLVLLVLRKKRI-----KEPLLIPEDDTRDNVF 699
660 AGVPLGLGLGIGI--LLFVLLILKYLILPYLLFFSSCSCKTKTPLLSSVDDVRNIC 717
700 YGEGEGGEDQ 711
718 YDEECGEGDQ 729

RESULT 13
Q7SZW3
ID Q7SZW3 PRELIMINARY; PRT; 677 AA.
AC Q7SZW3;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE SI:DZ71B9.1 (Novel protein similar to zebrafish epithelial cadherin 1 (Cdh1)) (Fragment)
DE SI:DZ71B9.1.
GN SI:DZ71B9.1.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OC NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Bates K.;
RL Submitted (Aug-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL929295; CA93436.1; -.
FT NON_TER 1
SQ SEQUENCE 677 AA; 74493 MW; AF895F3165EBA5F7 CRC64;

Query Match 28.5%; Score 1245; DB 13; Length 677;
Best Local Similarity 43.9%; Pred. No. 1.3e-78;
Matches 260; Conservative 101; Mismatches 221; Indels 10; Gaps 7;

QY 96 PP-SKRILRRHQRDMVWAPIVSPENGKGFPPORLNQLKSNKDRTKIFYITGADSP 154
DB 83 PPKSSTGLKRAKRGWIIPFVSFVFNKSGKPPFMRVLVQIKSDSASQYWAYKITGEGADDP 142
QY 155 EGVAFAVEKETGMLLNKPLDREETAKEFLFGHVSNGASVEDPMNT-SIIITDQNDHKPK 214
DB 143 KGIFKIDRLSGWINTVQQLDREKASYKLVAHATGVDVNIKEPFRITVTTDQNDKVP 202
QY 215 FTQDTFRGSLVGLPGTSMVQVATDDEDAIYVNGVAVSIHQSPKDPHDLMTIHR 274
DB 203 FTQNPFNANVPEALEKGEVFTVATDADDKENTDNADISYAIISQDPPSAKP2MFAINP 262
QY 275 STGTISVTSGLDREKPEYTLTIOATMDGSGTTTAVAVVEILDANDNAPMDPKYE 334
DB 263 VSGGHSVLETGLDREKQWRYTLVITATDNGEGSTTGATVITVNDSDNAPLEQTSHT 322
QY 335 AHVPENAVGHEVQRLTITDLPAPNSPAWRATYLYLIMGDDGPHFTITTHPENQGLITRK 394
DB 323 VSVPENQVGAQVANLPVTDGPEPESTANSTKYRVIGDGKGFNFVSTGPSLEGVITVK 382
QY 395 GLDFFAKQHTLYVEVTNEAPVLKPTSTATIVVHVEDVNEAPVFPVPSKVVEQEGIP 454
DB 383 PLDYETKQYILFVIVNDDKFGVSLPTSTATVTVNVEDVNEPPEFIPKEFISKPELDLP 442
QY 455 TCEPVCVVTAEADPKDE-NOKISYRILRDPAGWLAMPDPSQGVATVGLTLDREDEQVFNRI 513
DB 443 VGSNLVIYATDVTDEKQKITRYIDYDPGWLSI-TDSGQIQVKMTADRESSN-VKDGK 500

```

```

514 YEVWLVANDNG-SPTTGTGTLILLTLIDVNDHGPVPEPRQITICNQSPVRHVNLITDKOL 572
501 YKALILALDQDDVESPATGTGTVLIELQVNDNAPVINERTIKLCNRESALVLLSITDNDG 560
573 SPTSPFQAQLTDDSDIYWTAEVNEEGDVTYLSLKFLKQDVTYVHLSDHGNKEQLTV 632
561 PFAAGPVSFVEQDTSKWSVFNETHYLSLKQSRDLQGEYNVLRVTDREGESQENT 620
633 IPATYCDCHGHVETCPGPKGK-----FILPVLGAVALLFLLVLLVLRKK 680
621 IQASVCDCKGEAFQCTENQEAQGRPLFGVLGVLGGIL-LLLLLGLLSMFLRK 671

RESULT 14
Q7ZY6
ID Q7ZY6 PRELIMINARY; PRT; 690 AA.
AC Q7ZY6;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE SI:DZ18005.2 (Novel protein similar to E-cadherin (cdh1)) (Fragment)
DE SI:DZ18005.2.
GN SI:DZ18005.2.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OC NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Pelan S.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL929461; CAD60789.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
DR InterPro; IPR002126; Cadherin.
DR Pfam; PF00028; cadherin; 5.
DR SMART; SM00112; Ca; 4.
DR PROSITE; PS00232; CADHERIN_1; 2.
DR PROSITE; PS50268; CADHERIN_2; 4.
FT NON_TER 1
FT NON_TER 690
SQ SEQUENCE 690 AA; 75837 MW; 70C0AB808BF98357 CRC64;

Query Match 28.0%; Score 1221.5; DB 13; Length 690;
Best Local Similarity 39.7%; Pred. No. 6e-77;
Matches 278; Conservative 108; Mismatches 262; Indels 53; Gaps 12;

QY 48 GOALGK-VMGCGQEPALFSTNDNDFTVRNGETVQERRSLKERNPLK----- 94
DB 6 GERLCRWVESTCDGRGTIDFESTDQIEDLNMMDGLMRRSVTLHEGLKEFSVSAWDFSGK 65
QY 95 ----IFPSKILARRK-----RDWVAPIVSPENGKGFPPORLNQLKSNKDR 137
DB 66 KHTVFRVVRVDYHKDHDVNTLMTSSVQRCILFPPLTLHENGKGFPPRLVQIKSDSAS 125
QY 138 DTKIFYSIITGPGADSPPEGFAVEKETGMLLNKPLDREETAKEFLFGHVSNGASVED 197
DB 126 ETPMAYKITGEGADQPKGIFQIDRLSGWVSVTQQLDREKASYKLVAHATGVDVNIKEK 185
QY 198 PMNISIIVTQDNDHKPKFTQDTRFGSLVGLPGTSMVQVATDDEDAIYVNGVAVSI 257
DB 186 PFEFVTVADQNDKPLFTQNPFNANVPEALEKGEVFTVATDADDKENTDNADISYAI 245
QY 258 HSQEPKDPHDLMTIHRSTGTISVTSGLDREKPEYTLTIOATMDGSGTTTAVVE 317
DB 246 ISQDPPSPXPNMFALNPVSGGISMLETGLDREKQWRYTLVITATDNGEGSGSTTGTAVIT 305
QY 318 ILDANDNAPMFPQKYEAHVPENAVGHEVQRLTITDLPAPNSPAWRATYLYLIMGDDGPH 377
DB 306 VTDSNNNAPLFEQTLTYTVSVPNQVGAQVANLPVTDGPEPESTANSTKYRVIGDGKDF 365

```


This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 22:00:03 ; Search time 70 Seconds
(without alignments)

3346.170 Million cell updates/sec

Title: US-09-916-849A-1

Perfect score: 4369

Sequence: 1 MGUPRGPLASLLLLQVCWLQ.....NEWGSRFKLADWYGGGDD 829

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq_29Jan04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4369	100.0	829	5 ABG61897	Abg61897 Prostate
2	4369	100.0	829	5 ABB81476	Abb81476 Human P-c
3	4369	100.0	829	5 ABJ05598	Abj05598 Breast ca
4	4369	100.0	829	5 AAU97492	Aau97492 Human P-c
5	4369	100.0	829	5 AAM50864	Aam50864 Cadherin
6	4369	100.0	829	6 ABP58357	Abp58357 Human P-c
7	4369	100.0	829	7 ADC15497	Adc15497 Human bas
8	4369	100.0	829	7 ADD14190	Add14190 Human src
9	4361	99.8	829	5 ABP54683	Abp54683 Metastati
10	4361	99.8	829	6 ABR58670	Abr58670 Human can
11	4361	99.8	829	6 ABUS6434	Abus6434 Lung canc
12	4351	99.6	829	6 ABUS6670	Abus6670 Lung canc
13	2420	55.4	882	2 AAU93375	Aau93375 Wild-type
14	2420	55.4	882	3 AAB35730	Aab35730 Human E-c
15	2420	55.4	882	4 AAB73490	Aab73490 Human E-c
16	2420	55.4	882	5 ABG96293	Abg96293 Human ova
17	2420	55.4	882	5 AAU78051	Aau78051 Human E-c
18	2388	54.7	899	5 ABB81472	Abb81472 Mouse E-c
19	2348.5	53.8	878	2 AAR85487	Aar85487 Human E-c
20	2341.5	53.6	878	2 AAR55060	Aar55060 Sequence
21	2341.5	53.6	878	5 ABB81475	Abb81475 Human E-c
22	1714	39.2	906	3 AAU70741	Aay70741 Human N-c
23	1714	39.2	906	7 ADE55478	Ade55478 Human Pro
24	1714	39.2	906	7 ADE55482	Ade55482 Human Pro
25	1714	39.2	906	7 ADE55486	Ade55486 Human Pro

26	1714	39.2	906	7 ADE55490	Ades55490 Human Pro
27	1713	39.2	906	6 ABR58643	Abr58643 Human can
28	1708	39.1	906	4 ABG14316	Abg14316 Novel hum
29	1707	39.1	906	5 ABB81474	Abb81474 Human N-c
30	1707	39.1	906	6 ABR47406	Abr47406 Breast ca
31	1704.5	39.0	912	5 ABB81471	Abb81471 Chicken N
32	1697.5	38.9	906	5 ABB57233	Abb57233 Mouse isc
33	1693	38.8	906	7 ADE55488	Ades55488 Rat Prote
34	1693	38.8	906	7 ADE55476	Ades5476 Rat Prote
35	1693	38.8	906	7 ADE55484	Ades5484 Rat Prote
36	1693	38.8	906	7 ADE55480	Ades5480 Rat Prote
37	1662	38.0	747	7 ADD14035	Add14035 Human src
38	1638	37.5	824	7 ADB64361	Adb64361 Human pro
39	1621.5	37.1	916	2 AAW25658	Aaw25658 Human cad
40	1621.5	37.1	916	2 AAW13129	Aaw13129 Full leng
41	1621.5	37.1	916	7 ADD14110	Add14110 Human src
42	1422	32.5	848	5 ABG79689	Abg79689 Tumour in
43	1321.5	30.2	814	3 AAB24089	Aab24089 Human PRO
44	1321.5	30.2	814	4 ABG30224	Abg30224 Novel hum
45	1321.5	30.2	814	5 AAE26669	Aae26669 Human cad

ALIGNMENTS

RESULT 1

ABG61897

ID ABG61897 standard; protein; 829 AA.

XX

AC ABG61897;

XX

DT 15-AUG-2002 (first entry)

XX

DE Prostate cancer-associated protein #98.

XX

KW Prostate cancer; prostate tumour tissue; human; mammal; cytostatic.

XX

OS Mammalia.

XX

PN WC200230268-A2.

XX

PD 18-APR-2002.

XX

PF 12-OCT-2001; 2001WO-US032045.

XX

PR 13-OCT-2000; 2000US-00687576.

PR

PR 08-DEC-2000; 2000US-00733288.

PR

PR 08-DEC-2000; 2000US-00733742.

PR

PR 24-JAN-2001; 2001US-02639572.

PR

PR 16-MAR-2001; 2001US-0276791P.

PR

PR 16-MAR-2001; 2001US-0276888P.

PR

PR 06-APR-2001; 2001US-0281922P.

PR

PR 24-APR-2001; 2001US-0286214P.

PR

PR 30-APR-2001; 2001US-00847046.

PR

PR 04-MAY-2001; 2001US-0288589P.

XX

XX (EOSB-) EOS BIOTECHNOLOGY INC.

PA

Gish KC, Mack DH, Wilson KE, Afar D, Hevezi P;

XX

PI WPI: 2002-471335/50.

XX

DR N-PSDB; ABK92214.

XX

XX Detecting a prostate cancer-associated transcript in a cell in a patient,

PT useful for diagnosing prostate cancer (PC) or screening modulators of PC,

PT by determining if prostate cancer-associated genes are expressed in a

PT prostate tissue.

XX

XX Claim 27; Page 384-385; 436pp; English.

PS

XX The present invention relates to methods of detecting a prostate cancer-

CC associated transcript in a cell from a patient. The method comprises

CC contacting a biological sample from the patient with prostate cancer-

associated polynucleotides (designated PC genes) that selectively hybridise to a sequence that is at least 80% identical to them. The prostate cancer-associated polynucleotide sequences are differentially expressed in prostate tumour tissue or in prostate cancer and are derived from the tissues of various organisms such as humans or other mammals (e.g. mice, sheep and dogs). The methods of the invention are useful for diagnosing and treating prostate cancer in mammals. The prostate cancer-associated genes are useful for diagnosing or treating prostate cancer, as well as for identifying modulators of prostate cancer or agents that inhibit prostate cancer. The nucleic acid sequences are particularly useful in gene therapy, as a vaccine or in antisense applications.

ABG61800-ABG61944 represent prostate cancer-associated proteins

XX SQ Sequence 829 AA;

Query Match 100.0%; Score 4369; DB 5; Length 829;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPRGPLASILLQVCWLQCAASEPCRAVFAEVTLEAGGAEQBPQALGVFMGCPG 60

DB 1 MGLPRGPLASILLQVCWLQCAASEPCRAVFAEVTLEAGGAEQBPQALGVFMGCPG 60

QY 61 QEPALFSTDDNDFTVRGETVQERRSKERNPLKIPSKILRHKRDWVAVISPENG 120

DB 61 QEPALFSTDDNDFTVRGETVQERRSKERNPLKIPSKILRHKRDWVAVISPENG 120

QY 121 KGPPFQRLNQLSKNDRTKIFYSITGFGADSPPEGVFAVEKETGMLLNKPLDREEIAK 180

DB 121 KGPPFQRLNQLSKNDRTKIFYSITGFGADSPPEGVFAVEKETGMLLNKPLDREEIAK 180

QY 181 YELFGHAVSENGASVEDPMNISIVTQNDHKPKFTQDTRGSLGVLPGTSVMQVAT 240

DB 181 YELFGHAVSENGASVEDPMNISIVTQNDHKPKFTQDTRGSLGVLPGTSVMQVAT 240

QY 241 DEDDAIYVNGVAYS:HSQEPKDPHLMFTIHRSTGTSIVSSGLDREKVPYTLTIOA 300

DB 241 DEDDAIYVNGVAYS:HSQEPKDPHLMFTIHRSTGTSIVSSGLDREKVPYTLTIOA 300

QY 301 TMDGDSGTTTAVAVVILDANDNAPMFDQKYEAHVPENAVGHEVQLTVDLDAPNSP 360

DB 301 TMDGDSGTTTAVAVVILDANDNAPMFDQKYEAHVPENAVGHEVQLTVDLDAPNSP 360

QY 361 AWRATYILMGDDGDHFTIITHPESNOGILTRKGLDPEAKNOHTLYVEVNEAPFVLKL 420

DB 361 AWRATYILMGDDGDHFTIITHPESNOGILTRKGLDPEAKNOHTLYVEVNEAPFVLKL 420

QY 421 PTSTATIVHVEDYNEAFVFPVPSKVVEQGIPTGEPVCVYTAEDPDKENQKISYRLLR 480

DB 421 PTSTATIVHVEDYNEAFVFPVPSKVVEQGIPTGEPVCVYTAEDPDKENQKISYRLLR 480

QY 481 DPAGWLAMPDSGQVAVGLTDRDEQFVRNNIYEVWVLAMDNGSPPTTGTLTLTLD 540

DB 481 DPAGWLAMPDSGQVAVGLTDRDEQFVRNNIYEVWVLAMDNGSPPTTGTLTLTLD 540

QY 541 VNDHGVPPEPQITICNQSPVRHVLNITDKDLSPTSPFQAQLTDDSDIYWTAEVNEEGD 600

DB 541 VNDHGVPPEPQITICNQSPVRHVLNITDKDLSPTSPFQAQLTDDSDIYWTAEVNEEGD 600

QY 601 TVVLSLKKFLKQDYDYLHSLSDHGNKEQLTIRATVCDCHGVETCPGWKGGFLLPVL 660

DB 601 TVVLSLKKFLKQDYDYLHSLSDHGNKEQLTIRATVCDCHGVETCPGWKGGFLLPVL 660

QY 661 GAVLALLFLVLLVLLVLLVLLVLLVLLVLLVLLVLLVLLVLLVLLVLLVLLVLLVLL 720

DB 661 GAVLALLFLVLLVLLVLLVLLVLLVLLVLLVLLVLLVLLVLLVLLVLLVLLVLLVLL 720

QY 721 GLEARPEVVLNDVAPTIITPTMYRPRPANDPEIGNFIENLKAANTDTPAPPYDILLVF 780

DB 721 GLEARPEVVLNDVAPTIITPTMYRPRPANDPEIGNFIENLKAANTDTPAPPYDILLVF 780

QY 781 DYEGSGDAASLSLTSSASDQDQDYLVNENWGSFRFKKLADMYGGGEDD 829

DB 781 DYEGSGDAASLSLTSSASDQDQDYLVNENWGSFRFKKLADMYGGGEDD 829

RESULT 2

ABB81476

ID ABB81476 standard; protein; 829 AA.

AC ABB81476;

XX 30-AUG-2002 (first entry)

DT Human P-cadherin protein SEQ ID NO:50.

DE Cadherin; alpha-catenin; cancer; beta-catenin binding domain; melanoma;

XX o-catenin; colon cancer.

OS Homo sapiens.

XX US2002045591-A1.

XX 18-APR-2002.

XX 17-JUL-2001; 2001US-00905983.

XX 26-MAY-1998; 98IL-00124650.

PR 26-MAY-1999; 99US-00318633.

XX (GEIG/) GEIGER B.

PA (BENZ/) BEN-ZE'EV A.

PA (SADO/) SADOT E.

XX Geiger B, Ben-Ze'ev A, Sadot E;

PI WPI; 2002-499105/53.

DR N-PSDB; ABN89393.

XX New construct encoding soluble cytoplasmic portion of cadherin including beta catenin binding domain useful in treating cancer associated with high beta-catenin activity e.g. colon cancer and melanoma.

Example 3; Page 53-55; 102pp; English.

The present invention describes a pharmaceutical composition for treating cancer associated with abnormally high beta-catenin activity. The pharmaceutical composition comprises a gene therapy vehicle harbouring a polynucleotide that contains: (i) a nucleotide sequence encoding a soluble cytoplasmic portion of a cadherin which lacks a transmembrane portion and an extracellular portion of the cadherin, and includes a beta-catenin binding domain; and (b) an upstream promoter for directing expression of the soluble cytoplasmic portion of the cadherin in a mammalian cell. Also described is a pharmaceutical composition for treating cancer associated with abnormally high activity levels of beta-catenin comprising a gene therapy vehicle harbouring a polynucleotide that contains: (a) a nucleotide sequence encoding an o-catenin; and (b) an upstream promoter for directing expression of the o-catenin in a mammalian cell. The pharmaceutical compositions have cytoskeletal activity and can be used in the suppression of beta-catenin-mediated transactivation. They can be used for treating cancers associated with abnormally high activity levels of beta-catenin such as colon cancers and melanomas, by reducing these high activity levels of beta-catenin in mammalian cells. The present sequence represents human P-cadherin which is used in the exemplification of the present invention

XX SQ Sequence 829 AA;

Query Match 100.0%; Score 4369; DB 5; Length 829;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPRGPLASILLQVCWLQCAASEPCRAVFAEVTLEAGGAEQBPQALGVFMGCPG 60

DB 1 MGLPRGPLASILLQVCWLQCAASEPCRAVFAEVTLEAGGAEQBPQALGVFMGCPG 60

Qy 61 QEPALFSTDDNDFTVRNGETVQERRSLKERNPLKIFPSKILRRHKRDWVAPISVPENG 120
Db 61 QEPALFSTDDNDFTVRNGETVQERRSLKERNPLKIFPSKILRRHKRDWVAPISVPENG 120
Qy 121 KGPPFQRLNQLKSNKORDTKIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDRREIAK 180
Db 121 KGPPFQRLNQLKSNKORDTKIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDRREIAK 180
Qy 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTODTFRGSVLEGLVPGTSMQVAT 240
Db 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTODTFRGSVLEGLVPGTSMQVAT 240
Qy 241 DEDDAIYTYNGVAVSIHSEPKDPHDLMTTHRSTGTISVISGLDREKVPYTLTIOA 300
Db 241 DEDDAIYTYNGVAVSIHSEPKDPHDLMTTHRSTGTISVISGLDREKVPYTLTIOA 300
Qy 301 TMDGCGSTTTAVAVVEILLDANDNAPMFDPOKYEAHVPENAVGEVORLVTDLDPNSP 360
Db 301 TMDGCGSTTTAVAVVEILLDANDNAPMFDPOKYEAHVPENAVGEVORLVTDLDPNSP 360
Qy 361 AWRATYILMGDDGDHFTTTHPESNQGLTTRKGLDPEAKNQHTLYVEVTNEAPFVLK 420
Db 361 AWRATYILMGDDGDHFTTTHPESNQGLTTRKGLDPEAKNQHTLYVEVTNEAPFVLK 420
Qy 421 PTSTATIVHVEDVNEAPVFPVPSKVVEQEGIPTEGPEVCVYTAEDDPKKNQKISYRILR 480
Db 421 PTSTATIVHVEDVNEAPVFPVPSKVVEQEGIPTEGPEVCVYTAEDDPKKNQKISYRILR 480
Qy 481 DPAGWLAMPDPSGOVTAVGTLDRDEQFVRNNIYEVNVLAMNGSPPTTGTGTLTLLID 540
Db 481 DPAGWLAMPDPSGOVTAVGTLDRDEQFVRNNIYEVNVLAMNGSPPTTGTGTLTLLID 540
Qy 541 VNRHGPVPERQITICNQSPVRHVLNITDKDLSPTSPFOAQLTDDSDIYWTAEVNEEGD 600
Db 541 VNRHGPVPERQITICNQSPVRHVLNITDKDLSPTSPFOAQLTDDSDIYWTAEVNEEGD 600
Qy 601 TVVLSLKKFLKQDIYDVHLSLSDHGNKEQLTVIRATVCDCHGVETCPGPKWGGFILPVL 660
Db 601 TVVLSLKKFLKQDIYDVHLSLSDHGNKEQLTVIRATVCDCHGVETCPGPKWGGFILPVL 660
Qy 661 GAVLALLFLLLVLLVLRKKRIKEPILLPEDDTRDNVYFYGEGGEEQDYDITQLHR 720
Db 661 GAVLALLFLLLVLLVLRKKRIKEPILLPEDDTRDNVYFYGEGGEEQDYDITQLHR 720
Qy 721 GLEARPEVLNDVAPTIITPMYRPRPNDPEIGNFIENLKAANTDPTAPPYDTLLVF 780
Db 721 GLEARPEVLNDVAPTIITPMYRPRPNDPEIGNFIENLKAANTDPTAPPYDTLLVF 780
Qy 781 DYEGSGSDAASLSLTSASDQDDYDYLNEWGSFRFKLADMYGGGEDD 829
Db 781 DYEGSGSDAASLSLTSASDQDDYDYLNEWGSFRFKLADMYGGGEDD 829

RESULT 3
ID ABJ05598
AC ABJ05598 standard; protein; 829 AA.
XX
XX
XX
XX
DT 14-NOV-2002 (first entry)
DE Breast cancer-associated protein 63.
XX Breast cancer; breast cancer-associated gene sequence; drug development;
XX pharmacogenetics; biosensor development.
XX Unidentified.
XX
XX WO200259377-A2.
XX
XX 01-AUG-2002.
XX
XX 24-JAN-2002; 2002WO-US002242.
XX
XX

XX
PR
PR
PR
PR
PR
PR
XX
PA
XX
XX
XX
DR
XX
DR
XX
PT
PT
PT
XX
XX
XX
XX
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
XX
SQ

24-JAN-2001; 2001US-0263965P.
02-FEB-2001; 2001US-0265928P.
09-APR-2001; 2001US-00829472.
09-APR-2001; 2001US-0282698P.
04-MAY-2001; 2001US-0288590P.
29-MAY-2001; 2001US-0294443P.
(EOSB-) EOS BIOTECHNOLOGY INC.
Mack DH, Gish KC, Afar D;
WPI; 2002-583738/62.
N-PSDB; ABT07755.
Detecting a breast cancer-associated transcript in a patient's cell,
sample with a polynucleotide that selectively hybridizes with breast
cancer nucleic acids.
Disclosure; Page 405; 414pp; English.
The invention comprises a method of detecting a breast cancer-associated
transcript in a cell from a patient. The method of the invention involves
contacting a biological sample from the patient with a nucleotide that
hybridizes to one of the 69 breast cancer-associated gene sequences shown
in the specification. The method of the invention is useful in the
diagnosis or prognosis of breast cancer, and for detecting genes that are
up or down-regulated in breast cancer cells. Genes identified by the
method of the invention can be used in diagnostic purposes and also as
targets for screening for therapeutic compounds that modulate breast
cancer (e.g. hormones or antibodies). Identification of genes that are
over or under expressed in breast cancer can additionally provide high-
resolution, high-sensitivity datasets which can be used in the areas of
diagnostics, therapeutics, drug development, pharmacogenetics, protein
structure and biosensor development. Amino acid sequences ABJ05536 -
ABJ05604 represent the proteins encoded by the 69 breast cancer-
associated genes of the invention

Sequence 829 AA;

Query Match 100.0%; Score 4369; DB 5; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MGLPRGLASLLQLQVCLQCAASEPCRAVFEAEVTLSEAGGAEQEPGQALCKVFMGCPG 60
Db 1 MGLPRGLASLLQLQVCLQCAASEPCRAVFEAEVTLSEAGGAEQEPGQALCKVFMGCPG 60
Qy 61 QEPALFSTDDNDFTVRNGETVQERRSLKERNPLKIFPSKILRRHKRDWVAPISVPENG 120
Db 61 QEPALFSTDDNDFTVRNGETVQERRSLKERNPLKIFPSKILRRHKRDWVAPISVPENG 120
Qy 121 KGPPFQRLNQLKSNKORDTKIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDRREIAK 180
Db 121 KGPPFQRLNQLKSNKORDTKIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDRREIAK 180
Qy 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTODTFRGSVLEGLVPGTSMQVAT 240
Db 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTODTFRGSVLEGLVPGTSMQVAT 240
Qy 241 DEDDAIYTYNGVAVSIHSEPKDPHDLMTTHRSTGTISVISGLDREKVPYTLTIOA 300
Db 241 DEDDAIYTYNGVAVSIHSEPKDPHDLMTTHRSTGTISVISGLDREKVPYTLTIOA 300
Qy 301 TMDGCGSTTTAVAVVEILLDANDNAPMFDPOKYEAHVPENAVGEVORLVTDLDPNSP 360
Db 301 TMDGCGSTTTAVAVVEILLDANDNAPMFDPOKYEAHVPENAVGEVORLVTDLDPNSP 360
Qy 361 AWRATYILMGDDGDHFTTTHPESNQGLTTRKGLDPEAKNQHTLYVEVTNEAPFVLK 420
Db 361 AWRATYILMGDDGDHFTTTHPESNQGLTTRKGLDPEAKNQHTLYVEVTNEAPFVLK 420

QY 421 PTSTATIVHVEDVNEAPVFPVPPSKVVEVQEGITGEPVCVYTAEDPKENQKISYRILR 480
DB 421 PTSTATIVHVEDVNEAPVFPVPPSKVVEVQEGITGEPVCVYTAEDPKENQKISYRILR 480
QY 481 DPAGWLAMPDPSGGQVAVGTGLDREDEQFVRNRIYEVNVLAMDNGSPPTTGTGLLLTLID 540
DB 481 DPAGWLAMPDPSGGQVAVGTGLDREDEQFVRNRIYEVNVLAMDNGSPPTTGTGLLLTLID 540
QY 541 VNDHGPVPEPQIITICNOSPVRHVLNITDKLSHSTSPFOAQLTDDSDIYVTAEVNREGD 600
DB 541 VNDHGPVPEPQIITICNOSPVRHVLNITDKLSHSTSPFOAQLTDDSDIYVTAEVNREGD 600
QY 601 TVVLSLKKFLKQDITYDVHLSLSHDGNEQLTVIRATVCDCHGHVETCPGPKWGGFLLPVL 660
DB 601 TVVLSLKKFLKQDITYDVHLSLSHDGNEQLTVIRATVCDCHGHVETCPGPKWGGFLLPVL 660
QY 661 GAVLALLFLLVLLVLRKKKIKKEPILLPDDTRDNVFIYGEEGGEEDQDITQLHR 720
DB 661 GAVLALLFLLVLLVLRKKKIKKEPILLPDDTRDNVFIYGEEGGEEDQDITQLHR 720
QY 721 GLEARPEVLRNDVAPTIIPTPMYRPRPNDPEIGNFIENLKAANTDPTAPPYDITLLVF 780
DB 721 GLEARPEVLRNDVAPTIIPTPMYRPRPNDPEIGNFIENLKAANTDPTAPPYDITLLVF 780
QY 781 DYEGSGDAASLSLTSASDQDQDYDLNWSGRFCKLADMYGGGEDD 829
DB 781 DYEGSGDAASLSLTSASDQDQDYDLNWSGRFCKLADMYGGGEDD 829

RESULT 4

AAU97492
ID AAU97492 standard; protein; 829 AA.

AC AAU97492;

DT 13-AUG-2002 (first entry)

XX Human P-cadherin (placental cadherin) protein.

DE Human; P-cadherin; placental cadherin; solid surface; vascular graft;
KW cell-to-cell cohesion; endothelial cell-coated surface;
KW hydrodynamic shear; shaking container; continuous harvest system;
KW laminar shear device.

XX Homo sapiens.

XX WO200231121-A2.

XX 18-APR-2002.

XX 15-OCT-2001; 2001WO-US032030.

XX 13-OCT-2000; 2000US-0241216P.

XX 27-OCT-2000; 2000US-0243693P.

XX 11-OCT-2001; 2001US-00975723.

XX (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.

XX Nackman G, Poty RA;

XX WPI; 2002-435443/46.

XX N-PSDB; ASK52038.

XX Populating solid surface especially biomedical devices with vascular
XX endothelial cells by increasing cell-to-cell cohesion useful to heal
XX vascular grafts.

XX Disclosure; Page 9-10; 38pp; English.

XX The present invention relates to a new method of populating a solid
XX surface with cells. The method of the invention involves increasing the
XX cell-to-cell cohesion of the cells. The method is useful for populating a
XX solid surface with human and nonhuman cells. The invention is also useful

CC for achieving better utilisation of vascular grafts and also in any
CC system that employs endothelial cell-coated surface susceptible to
CC hydrodynamic shear, such as shaking container coated with endothelial
CC cells, continuous harvest systems for harvesting the products of
CC endothelial cells on solid surface, and laminar shear devices. The
CC present amino acid sequence represents the human P-cadherin (placental
CC cadherin) protein of the invention

XX Sequence 829 AA;

Query Match 100.0%; Score 4369; DB 5; Length 829;
Best Local Similarity 100.0%; Pred.No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPRGPLASLLLLQVLCQAASPCRAVPREAVTLEAGAGQEPGQALGKVFMCQPG 60
DB 1 MGLPRGPLASLLLLQVLCQAASPCRAVPREAVTLEAGAGQEPGQALGKVFMCQPG 60
QY 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDVWVAPISVPENG 120
DB 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDVWVAPISVPENG 120
QY 121 KQFFPQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWLNLKPLDREIAK 180
DB 121 KQFFPQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWLNLKPLDREIAK 180
QY 181 YELFGHVAISENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSVLEGVLPCTSYMQVTAT 240
DB 181 YELFGHVAISENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSVLEGVLPCTSYMQVTAT 240
QY 241 DEDDAIYNGVWVAYSIHSQEPKOPHDLMTFTHSTGTISVISSGLDREKPEYTLTIOA 300
DB 241 DEDDAIYNGVWVAYSIHSQEPKOPHDLMTFTHSTGTISVISSGLDREKPEYTLTIOA 300
QY 301 TDMGDGSTTTAVAVVEILDANDNAPMFDPOKYBAHVPENAVGHVORLTITDLDAPNSP 360
DB 301 TDMGDGSTTTAVAVVEILDANDNAPMFDPOKYBAHVPENAVGHVORLTITDLDAPNSP 360
QY 361 AWRYLIMGDDGDHFTTTHPSNQILITRGLDPEAKNQHTLVYEVNTEAPFVLKL 420
DB 361 AWRYLIMGDDGDHFTTTHPSNQILITRGLDPEAKNQHTLVYEVNTEAPFVLKL 420
QY 421 PTSTATIVHVEDVNEAPVFPVPPSKVVEVQEGITGEPVCVYTAEDPKENQKISYRILR 480
DB 421 PTSTATIVHVEDVNEAPVFPVPPSKVVEVQEGITGEPVCVYTAEDPKENQKISYRILR 480
QY 481 DPAGWLAMPDPSGGQVAVGTGLDREDEQFVRNRIYEVNVLAMDNGSPPTTGTGLLLTLID 540
DB 481 DPAGWLAMPDPSGGQVAVGTGLDREDEQFVRNRIYEVNVLAMDNGSPPTTGTGLLLTLID 540
QY 541 VNDHGPVPEPQIITICNOSPVRHVLNITDKLSHSTSPFOAQLTDDSDIYVTAEVNREGD 600
DB 541 VNDHGPVPEPQIITICNOSPVRHVLNITDKLSHSTSPFOAQLTDDSDIYVTAEVNREGD 600
QY 601 TVVLSLKKFLKQDITYDVHLSLSHDGNEQLTVIRATVCDCHGHVETCPGPKWGGFLLPVL 660
DB 601 TVVLSLKKFLKQDITYDVHLSLSHDGNEQLTVIRATVCDCHGHVETCPGPKWGGFLLPVL 660
QY 661 GAVLALLFLLVLLVLRKKKIKKEPILLPDDTRDNVFIYGEEGGEEDQDITQLHR 720
DB 661 GAVLALLFLLVLLVLRKKKIKKEPILLPDDTRDNVFIYGEEGGEEDQDITQLHR 720
QY 721 GLEARPEVLRNDVAPTIIPTPMYRPRPNDPEIGNFIENLKAANTDPTAPPYDITLLVF 780
DB 721 GLEARPEVLRNDVAPTIIPTPMYRPRPNDPEIGNFIENLKAANTDPTAPPYDITLLVF 780
QY 781 DYEGSGDAASLSLTSASDQDQDYDLNWSGRFCKLADMYGGGEDD 829
DB 781 DYEGSGDAASLSLTSASDQDQDYDLNWSGRFCKLADMYGGGEDD 829

RESULT 5
AAMS0864

PR 31-MAY-2001; 2001US-0294225P.
 XX (CHIR) CHIRON CORP.
 PA Reinhard C, Klinger J, Jefferson AB, Escobedo J, Randazo F;
 PI Winter J, Goodson R;
 XX N-PSDB; ABZ24736.
 DR WPI: 2003-140501/13.
 XX Inhibiting migration and proliferation of P-cadherin expressing cancer.
 PT for treating cancer, especially digestive cancer, characterized by
 PT overexpression of P-cadherin, involves administering a P-cadherin
 PT antagonist.
 XX Disclosure; Page 10-11; 129pp; English.
 XX
 CC The present sequence is the protein sequence for human P-cadherin
 CC (placental cadherin), a calcium-dependent cellular adhesion protein. The
 CC invention provides methods of treating or diagnosing cancers involving P-
 CC cadherin expression using ligands that target P-cadherin, especially
 CC human anti-P-cadherin antibodies. A claimed method of treating a cancer
 CC characterised by the overexpression and/or upregulation of P-cadherin
 CC comprises the administration of a P-cadherin antagonist, optionally
 CC conjugated to a therapeutic agent. The migration, adhesion and/or
 CC proliferation of the cancer is inhibited, and the method is especially
 CC useful for treating or preventing a digestive cancer such as colon or
 CC colorectal cancer. The antagonists may be an anti-P-cadherin antibody or
 CC its fragment, a ribozyme or antisense oligonucleotide. A transgenic
 CC animal that expresses a recombinant (human) antibody that specifically
 CC binds P-cadherin is also claimed. A method of determining the presence of
 CC cancer involves determining the expression level of P-cadherin in a cell
 CC sample. Also provided are screens for identifying anti-P-cadherin
 CC antibodies have therapeutic activity
 XX Sequence 829 AA;
 SQ

Query Match 100.0%; Score 4369; DB 6; Length 829;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPRGPLASLLLLQVLCWQCAASEPCRAVFREAVTLEAGGAEQEPGQALGVFMGCPG 60
 DB 1 MGLPRGPLASLLLLQVLCWQCAASEPCRAVFREAVTLEAGGAEQEPGQALGVFMGCPG 60
 QY 61 QEPALFSTDDNDFTVRNGETVQERRSLKERNPLKIPPSKILRRHKRDWVAVISVPENG 120
 DB 61 QEPALFSTDDNDFTVRNGETVQERRSLKERNPLKIPPSKILRRHKRDWVAVISVPENG 120
 QY 121 KGPPQRLNOLKSNKRDITKIFYSTIGPGADSPPEGFAVEKETGMLLNKPLDRREIAK 180
 DB 121 KGPPQRLNOLKSNKRDITKIFYSTIGPGADSPPEGFAVEKETGMLLNKPLDRREIAK 180
 QY 181 YELFGHAVSENGASVEDPMNISIITVDNDHKPKFTQDTFRGSLVGLPFGTSVMQVAT 240
 DB 181 YELFGHAVSENGASVEDPMNISIITVDNDHKPKFTQDTFRGSLVGLPFGTSVMQVAT 240
 QY 241 DEDDAIYTYNGVAVSIHSGEPKDPHLMETIHSTGTISVISGLDREKVPYTLTIOA 300
 DB 241 DEDDAIYTYNGVAVSIHSGEPKDPHLMETIHSTGTISVISGLDREKVPYTLTIOA 300
 QY 301 TDMGDSGTTTAVAVVEILDANDNAPFDPQKYEAHPENAVGHEVORLTVTDLDPNSP 360
 DB 301 TDMGDSGTTTAVAVVEILDANDNAPFDPQKYEAHPENAVGHEVORLTVTDLDPNSP 360
 QY 361 AWRTATYLMGDDGCHTITTHPSNQGILTRKGLDFEAKNHTLYVETNEAPFVKL 420
 DB 361 AWRTATYLMGDDGCHTITTHPSNQGILTRKGLDFEAKNHTLYVETNEAPFVKL 420
 QY 421 PTSTATIVHVEDVNEAPFVPPSKVVEVQEGIPTGFPVCVYTAEDPDKENQISYRLR 480
 DB 421 PTSTATIVHVEDVNEAPFVPPSKVVEVQEGIPTGFPVCVYTAEDPDKENQISYRLR 480

QY 481 DPAGWLAMPDPSGGQVAVGTLDREDEQFVRNNIYEVWVYAMDNQSPPTTGTGTLTLLID 540
 DB 481 DPAGWLAMPDPSGGQVAVGTLDREDEQFVRNNIYEVWVYAMDNQSPPTTGTGTLTLLID 540
 QY 541 VNDHGVPPEPRQITICNOSPVRHVLNITDKDLSHTSPFQALTDSDDIYWTAEVNEGD 600
 DB 541 VNDHGVPPEPRQITICNOSPVRHVLNITDKDLSHTSPFQALTDSDDIYWTAEVNEGD 600
 QY 601 TVVLSLKKFLKQDITVDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFILLVL 660
 DB 601 TVVLSLKKFLKQDITVDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFILLVL 660
 QY 661 GAVLALLFLLVLLVLLVLRKKIKIPEPLLPEDDTRDNVFFYGGEGGEDDQDITQIHR 720
 DB 661 GAVLALLFLLVLLVLLVLRKKIKIPEPLLPEDDTRDNVFFYGGEGGEDDQDITQIHR 720
 QY 721 GLEARPEVVLNDVAPTIIPTPMYRPRANDEIGNFIENLKAANTDPTAPPYDTLLVF 780
 DB 721 GLEARPEVVLNDVAPTIIPTPMYRPRANDEIGNFIENLKAANTDPTAPPYDTLLVF 780
 QY 781 DYEGSGSDAASLSLTSSASDQDQDYDYLNEWGSRFKKLADMYGGGDD 829
 DB 781 DYEGSGSDAASLSLTSSASDQDQDYDYLNEWGSRFKKLADMYGGGDD 829

RESULT 7
 ADC15497
 ID ADC15497 standard; protein; 829 AA.
 XX AC ADC15497;
 XX AC ADC15497;
 DT 18-DEC-2003 (first entry)
 XX Human basal cell marker Cadherin-3.
 DE Human;
 KW Human; basal cell marker; cadherin-3; tumour; P-cadherin;
 KW matrix metalloproteinase 14;
 KW cadherin EGF LAG seven-pass G-type receptor 2;
 KW EGF-like Domain Multiple 2; mammary lactation gland; breast tumour;
 KW cancer; cytostatic.
 XX Homo sapiens.
 OS US2003086934-A1.
 EN 08-MAY-2003.
 PD 26-JUL-2001; 2001US-00916849.
 PF 26-JUL-2000; 2000US-0220967P.
 PR (BOTS/) BOTSTEIN D.
 PA (BROW/) BROWN P O.
 PA (PERO/) PEROU C M.
 PA (RING/) RING B.
 PA (ROSS/) ROSS D.
 PA (SEIT/) SEITZ R.
 PA (VRIJ/) VAN DE RIJN J M.
 XX Botstein D, Brown PO, Perou CM, Ring B, Ross D, Seitz R;
 PI Van De Riijn JM;
 DR WPI: 2003-786931/74.
 XX Classifying breast tumor by detecting expression or activity of gene
 PT encoding cadherin 3, matrix metalloproteinase 14, cadherin EGF LAG seven-
 PT pass G-type receptor 2 in tumor sample.
 XX Claim 1; SEQ ID NO 1; 89pp; English.
 XX The invention relates to classifying (M1) a tumour comprising detecting
 CC expression or activity of a gene encoding cadherin 3 or P-cadherin
 CC , matrix metalloproteinase 14 , or cadherin EGF LAG seven-pass G-type

receptor 2 or EGF-like domain, Multiple 2 (P1-3) or at least two genes chosen from gene encoding (P1), (P2) or (P3) in the tumour sample, and classifying tumour as belonging to a tumour subclass based on results of detecting step. Also included are testing, or predictive information based P1-3 and providing diagnostic, prognostic, or predictive information based on the detecting step, stratifying the subject for a clinical trial based on the detecting step, or selecting a treatment based on the detecting step, an antibody that specifically binds to an epitope found in (P1), (P2) or (P3) (where the antibody recognises basal cells in normal mammary lactation glands), methods of testing a compound or a combination of compounds for activity against tumours. The method (M1) is useful for classifying a tumour (breast tumour), as belonging to a tumour subclass such as basal tumour subclass. The antibody is useful for treating a subject in need of treatment for cancer. The tumour is a breast tumour, and where the method further involves identifying the tumour as belonging to a basal tumour subclass. The antibody is conjugated with a toxic molecule. The present sequence is human Cadherin-3.

Sequence 829 AA;

Query Match		100.0%;	Score 4369;	DB 7;	Length 829;
Best Local Similarity		100.0%;	Pred. No. 0;		
Matches: 829;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MGLPRGLASLLQLVCWLCQAASEPCRAVFREAEVTLAAGAEQEPGQALGVFMGCPG	60		
Db	1	MGLPRGLASLLQLVCWLCQAASEPCRAVFREAEVTLAAGAEQEPGQALGVFMGCPG	60		
Qy	61	QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIPPSKILRRHKRDWVAIPSPENG	120		
Db	61	QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIPPSKILRRHKRDWVAIPSPENG	120		
Qy	121	KGPPFQRLNQLKSNKRDTKIFYSITGPGADSPGEGVFAVEKETGMLLNKPLDREIAK	180		
Db	121	KGPPFQRLNQLKSNKRDTKIFYSITGPGADSPGEGVFAVEKETGMLLNKPLDREIAK	180		
Qy	181	YELFGHAVSENGASVEDPMNISIIITVDQNDHKPKFTQDTFRGSLVGLPSTVMQVAT	240		
Db	181	YELFGHAVSENGASVEDPMNISIIITVDQNDHKPKFTQDTFRGSLVGLPSTVMQVAT	240		
Qy	241	DEDAIYTVNGVAYSHSQEPKDPHLMFTIHRSTGTSIVISGGLDREKVPYTTIQA	300		
Db	241	DEDAIYTVNGVAYSHSQEPKDPHLMFTIHRSTGTSIVISGGLDREKVPYTTIQA	300		
Qy	301	TMDGSGSTTAVAVVEILDANDNAPFDPQKYEAHVPENAVGVHVEQRLTVTDLDPNSP	360		
Db	301	TMDGSGSTTAVAVVEILDANDNAPFDPQKYEAHVPENAVGVHVEQRLTVTDLDPNSP	360		
Qy	361	AWRATYILMGDDGDHPTIITHPESNGQILITTRKGLDFAKNQHTLYVEVTNEAPVVLK	420		
Db	361	AWRATYILMGDDGDHPTIITHPESNGQILITTRKGLDFAKNQHTLYVEVTNEAPVVLK	420		
Qy	421	PTSTATIVHVEDVNEAPVFPVPSKVVEQEGPTGEPVCVTAEDPDKENQKISYRIAR	480		
Db	421	PTSTATIVHVEDVNEAPVFPVPSKVVEQEGPTGEPVCVTAEDPDKENQKISYRIAR	480		
Qy	481	DPAGWLAMPDSQGVTAAGTLDREDEQFVRNNIYEVVWVLMANDNGSPPTTGTGLTLTLD	540		
Db	481	DPAGWLAMPDSQGVTAAGTLDREDEQFVRNNIYEVVWVLMANDNGSPPTTGTGLTLTLD	540		
Qy	541	VNDHGPVPEPQITICNQSPRVHVNITDKLSPTSPFOAQLTDSDIYWTAEVNEEGD	600		
Db	541	VNDHGPVPEPQITICNQSPRVHVNITDKLSPTSPFOAQLTDSDIYWTAEVNEEGD	600		
Qy	601	TVVLSLKFKLQDITVDVHLSLSHGHKNEQLTIVIRATVCDCHGVETCPGPKWGFFILPVL	660		
Db	601	TVVLSLKFKLQDITVDVHLSLSHGHKNEQLTIVIRATVCDCHGVETCPGPKWGFFILPVL	660		
Qy	661	GAVLALLFLLVLLVLLVLRKKRKKEPLLDPEDDTRDNVFFYEGEGGDEDDQDYITQLHR	720		
Db	661	GAVLALLFLLVLLVLLVLRKKRKKEPLLDPEDDTRDNVFFYEGEGGDEDDQDYITQLHR	720		
Qy	721	GLEARPEVLRNDVAPTIPTMYRPRPANPDEIGNFIENLKAANTDPTAPPYDTLLVF	780		

Db	721	GLEARPEVLRNDVAPTIPTMYRPRPANPDEIGNFIENLKAANTDPTAPPYDTLLVF	780		
Qy	781	DYEGSGSDAASISSTSSASDQDQDYDLNWSRPFKKLADMYGGEDD	829		
Db	781	DYEGSGSDAASISSTSSASDQDQDYDLNWSRPFKKLADMYGGEDD	829		
RESULT 8					
ID	ADD14190	standard; protein; 829 AA.			
XX	ADD14190;				
DT	01-JAN-2004	(first entry)			
DE	Human src biomarker polypeptide	SEQ ID NO:379.			
KW	predictor set; protein tyrosine kinase activity modulator;				
KW	protein tyrosine kinase pathway; protein tyrosine kinase; cytostatic;				
XX	Gene therapy; drug sensitivity; genetic profile; cancer; human.				
OS	Homo sapiens.				
XX	WO2003062395-A2.				
PD	31-JUL-2003.				
PF	17-JAN-2003; 2003WO-US001981.				
PR	18-JAN-2002; 2002US-0350061P.				
XX	(BRIM) BRISTOL-MYERS SQUIBB CO.				
PI	Huang F, Fairchild CR, Lee FY, Shaw P;				
DR	WPI; 2003-636735/60.				
XX	N-PSDB; ADD14795.				
PT	New polynucleotides and polypeptides for predicting the activity of				
PT	compounds that interact with protein tyrosine kinases and/or protein				
XX	tyrosine kinase pathways.				
PS	Claim 10; SEQ ID NO 379; 139pp; English.				
CC	The present invention describes a predictor set comprising a plurality of				
CC	polynucleotides or polypeptides whose expression pattern is predictive of				
CC	the response of cells to treatment with a compound that modulates protein				
CC	tyrosine kinase activity or members of the protein tyrosine kinase				
CC	pathway. Also described: (1) predicting whether a compound is capable of				
CC	modulating the activity of cells, comprising obtaining a sample of cells,				
CC	determining whether the cells express a plurality of markers, and				
CC	correlating the expression of the markers to the compound's ability to				
CC	modulate the activity of the cells; (2) a plurality of cell lines for				
CC	identifying polynucleotides and polypeptides whose expression levels				
CC	correlate with compound sensitivity or resistance of cells associated				
CC	with a disease state; and (3) identifying polynucleotides and				
CC	polypeptides that predict compound sensitivity or resistance of cells				
CC	associated with a disease state, comprising subjecting the plurality of				
CC	cell lines to one or more compounds, analysing the expression pattern of				
CC	a microarray of polynucleotides or polypeptides, and selecting				
CC	polynucleotides or polypeptides that predict the sensitivity or				
CC	resistance of cells associated with a disease state by using the				
CC	expression pattern of the microarray. The polynucleotides and				
CC	polypeptides have cytostatic activities, and can be used in gene therapy.				
CC	The polynucleotides and polypeptides are useful in predicting the				
CC	activity of compounds that interact with protein tyrosine kinases and/or				
CC	protein tyrosine kinase pathways. These may be used in determining drug				
CC	sensitivity in patients to allow the development of individualized				
CC	genetic profiles which aid in treating diseases and disorders (e.g.				
CC	cancer) based on patient response at a molecular level. The present				
CC	sequence is used in the exemplification of the present invention.				

SQ		Sequence 829 AA;	
Query Match		100.0%; Score 4369; DB 7; Length 829;	
Best Local Similarity		100.0%; Pred. No. 0;	
Matches 829; Conservative		0; Mismatches 0; Indels 0; Gaps 0;	
Qy	1	MGLPRGPLASLLLLQVWLQCAASEPCRAVPREAVTLEAGGAEQEPQALGKVFMC	60
Db	1	MGLPRGPLASLLLLQVWLQCAASEPCRAVPREAVTLEAGGAEQEPQALGKVFMC	60
Qy	61	QEPALFSTNDNDFTVRNGETVQERSLKERNPLKIFPSKRILRRHKRDWVAPISVP	120
Db	61	QEPALFSTNDNDFTVRNGETVQERSLKERNPLKIFPSKRILRRHKRDWVAPISVP	120
Qy	121	KGPPFQRLNQLKSNKDRDTKIFYSTGPGADSPPEGVFAVEKETGMLLNKPLDRE	180
Db	121	KGPPFQRLNQLKSNKDRDTKIFYSTGPGADSPPEGVFAVEKETGMLLNKPLDRE	180
Qy	181	YELFGHVSNGASVEDPMNISIIITDQNDHKPKFTQDTPRGSVLEGLPGETSV	240
Db	181	YELFGHVSNGASVEDPMNISIIITDQNDHKPKFTQDTPRGSVLEGLPGETSV	240
Qy	241	DEDDAIYTVNGVVAISYHSQEPKDPHDLMTIHRSTGTISVSSGLDREKVP	300
Db	241	DEDDAIYTVNGVVAISYHSQEPKDPHDLMTIHRSTGTISVSSGLDREKVP	300
Qy	301	TMGDGSGSTTTAVAVVEILDANDNAPMFPDQKYEAHVPENAVGHVORLT	360
Db	301	TMGDGSGSTTTAVAVVEILDANDNAPMFPDQKYEAHVPENAVGHVORLT	360
Qy	361	AWRATYVINGGDDGDFHTITTHPESNQILTRKGLDFEAKNQHTLYVEVTNE	420
Db	361	AWRATYVINGGDDGDFHTITTHPESNQILTRKGLDFEAKNQHTLYVEVTNE	420
RESULT 9		ABP54683	
ID	ABP54683	standard; protein; 829 AA.	
XX	XX	ABP54683;	
AC	ABP54683;		
DT	30-DEC-2002	(first entry)	
DE	Metastatic colorectal cancer-associated	polypeptide.	
XX	XX		
KW		Colorectal cancer; metastasis; differential expression; cytostatic;	
XX		diagnosis; gene therapy; vaccine.	
OS		Homo sapiens.	
XX		WO200268677-A2.	
PN		06-SEP-2002.	
XX		27-FEB-2002; 2002WO-US006001.	
PF		27-FEB-2001; 2001US-0272206P.	
XX		02-APR-2001; 2001US-0281149P.	
PR		17-APR-2001; 2001US-0284555P.	
XX		(EOSB-) EOS BIOTECHNOLOGY INC.	
PA		(UYCA-) UNIV CASE WESTERN RESERVE.	
XX		Mack DH, Markowitz SD;	
PI		WPI: 2002-698677/75.	
XX		N-PSDB; A8081547.	
DR		New genes that are up- or down-regulated in colorectal cancer, useful for	
XX		diagnosing colorectal cancer in a subject, or for identifying modulators	
PT		of colorectal cancer-associated proteins and genes for treating	
PT		colorectal cancer.	
XX		Claim 8; Page 245; 260pp; English.	
PS		The present sequence is the protein sequence of a human polypeptide	
XX		encoded by a gene that exhibits increased expression in colon cancer-	
CC		derived metastases compared to normal colon tissue. It is an example of	
CC		claimed polypeptides that are encoded by genes which are differentially	
CC		expressed in metastatic colorectal cancer cells. Such polypeptides are	
CC		useful in diagnostic and prognostic assays, for raising antibodies useful	
CC		e.g. in immunotherapy, and in screening for modulator compounds of	
CC		therapeutic value	
XX		Sequence 829 AA;	
SQ		Query Match 99.8%; Score 4361; DB 5; Length 829;	
Best Local Similarity		99.9%; Pred. No. 0;	
Matches 828; Conservative		0; Mismatches 1; Indels 0; Gaps 0;	
Qy	1	MGLPRGPLASLLLLQVWLQCAASEPCRAVPREAVTLEAGGAEQEPQALGKVFMC	60
Db	1	MGLPRGPLASLLLLQVWLQCAASEPCRAVPREAVTLEAGGAEQEPQALGKVFMC	60
Qy	61	QEPALFSTNDNDFTVRNGETVQERSLKERNPLKIFPSKRILRRHKRDWVAPISVP	120
Db	61	QEPALFSTNDNDFTVRNGETVQERSLKERNPLKIFPSKRILRRHKRDWVAPISVP	120
Qy	121	KGPPFQRLNQLKSNKDRDTKIFYSTGPGADSPPEGVFAVEKETGMLLNKPLDRE	180
Db	121	KGPPFQRLNQLKSNKDRDTKIFYSTGPGADSPPEGVFAVEKETGMLLNKPLDRE	180
Qy	181	YELFGHVSNGASVEDPMNISIIITDQNDHKPKFTQDTPRGSVLEGLPGETSV	240
Db	181	YELFGHVSNGASVEDPMNISIIITDQNDHKPKFTQDTPRGSVLEGLPGETSV	240
Qy	241	DEDDAIYTVNGVVAISYHSQEPKDPHDLMTIHRSTGTISVSSGLDREKVP	300
Db	241	DEDDAIYTVNGVVAISYHSQEPKDPHDLMTIHRSTGTISVSSGLDREKVP	300
Qy	301	TMGDGSGSTTTAVAVVEILDANDNAPMFPDQKYEAHVPENAVGHVORLT	360
Db	301	TMGDGSGSTTTAVAVVEILDANDNAPMFPDQKYEAHVPENAVGHVORLT	360
Qy	361	AWRATYVINGGDDGDFHTITTHPESNQILTRKGLDFEAKNQHTLYVEVTNE	420
Db	361	AWRATYVINGGDDGDFHTITTHPESNQILTRKGLDFEAKNQHTLYVEVTNE	420

QY 421 PTSTATIVHVEDVNEAPVFPVPSKVVEVQIGIPTGEPVVCVYTAEDDPKENOKISYRILR 480
 DB 421 PTSTATIVHVEDVNEAPVFPVPSKVVEVQIGIPTGEPVVCVYTAEDDPKENOKISYRILR 480
 QY 481 DPAGWLAMPDPSGQVAVGLDREDEQFVRNNIYEVNVLAMNDSPTTGTGILLTLID 540
 DB 481 DPAGWLAMPDPSGQVAVGLDREDEQFVRNNIYEVNVLAMNDSPTTGTGILLTLID 540
 QY 541 VNDHGPPPEPQITICNOSPVRHVNLITDKDLSPTSPFOAQLTDDSDIYVTAEVNEEGD 600
 DB 541 VNDHGPPPEPQITICNOSPVRHVNLITDKDLSPTSPFOAQLTDDSDIYVTAEVNEEGD 600
 QY 601 TVVLSLKKFLKQDVTYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFILLPV 660
 DB 601 TVVLSLKKFLKQDVTYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFILLPV 660
 QY 661 GAVLALLFLLLVLLVLRKKIKEPILLPEDDTRDNVYVYGGEGGEEDQDITQLHR 720
 DB 661 GAVLALLFLLLVLLVLRKKIKEPILLPEDDTRDNVYVYGGEGGEEDQDITQLHR 720
 QY 721 GLEARPEVLRNDVAPTIPTPMYRPRPANDPEIGNFIIENLKAANTDPTAPPYDTILLVF 780
 DB 721 GLEARPEVLRNDVAPTIPTPMYRPRPANDPEIGNFIIENLKAANTDPTAPPYDTILLVF 780
 QY 781 DYEGSGSDAASLSLTSASDQDDYDYNWGSRPKKIADMTGGGDD 829
 DB 781 DYEGSGSDAASLSLTSASDQDDYDYNWGSRPKKIADMTGGGDD 829

RESULT 10

ABR58670
 ID ABR58670 standard; protein; 829 AA.
 XX AC ABR58670;
 XX DT 09-JUL-2003 (first entry)
 XX DE Human cancer related protein SEQ ID NO:327.
 XX KW Human; cancer; diagnosis; screening; modulator; leukaemia; ischaemia;
 XX KW heart disease; atherosclerosis; endometriosis.
 XX OS Homo sapiens.
 XX PN W02003025138-A2.
 XX PD 27-MAR-2003.
 XX PF 17-SEP-2002; 2002WO-US029560.
 XX PR 17-SEP-2001; 2001US-0323469P.
 XX PR 20-SEP-2001; 2001US-0323887P.
 XX PR 13-NOV-2001; 2001US-0350666P.
 XX PR 08-FEB-2002; 2002US-0355145P.
 XX PR 08-FEB-2002; 2002US-0355257P.
 XX PR 12-APR-2002; 2002US-0372246P.
 XX PA (EOSB-) EOS BIOTECHNOLOGY INC.
 XX PI Afar D, Aziz N, Gish KC, Hevezi PA, Mack DH, Wilson KB;
 XX PI Zlotnik A;
 XX WPI; 2003-354600/33.
 XX DR N-PSDB; ACC72821.
 XX PT New genes that are up-regulated or down-regulated in cancers, useful as
 XX PT markers for diagnosing e.g. cancer, ischemia or heart diseases, or as
 XX PT therapeutic targets for screening drugs for treating these diseases.
 XX PS Claim 12; Page 757; 767pp; English.

CC The present invention describes an isolated nucleic acid molecule, which
 CC comprises the sequence of any of the genes that are up-regulated or down-

CC regulated in specific cancers (e.g. about 1031 genes up-regulated in
 CC acute lymphocytic leukemia).. ACC72641 to ACC72860 represent cancer
 CC related gene nucleotide sequences which encode the proteins given in
 CC ABR58521 to ABR58709. Also described: (1) determining the presence or
 CC absence of a nucleic acid cell in a patient; (2) an expression vector
 CC comprising a nucleic acid molecule described above; (3) a host cell
 CC comprising the vector; (4) an isolated polypeptide, which is encoded by
 CC the nucleic acid; (5) an antibody that specifically binds the polypeptide
 CC of (4); (6) specifically targeting a compound to a pathological cell in a
 CC patient by administering to the patient the antibody above; and (7) a
 CC drug screening assay. The nucleic acid is useful as diagnostic markers or
 CC therapeutic targets. In particular, the nucleic acid is useful for
 CC diagnosing a pathology, e.g. cancer (e.g. cancer of the bone marrow,
 CC bladder, brain, breast, cervix, colon/rectum, kidney, lung, ovary,
 CC pancreas, prostate, skin and uterus) wounds, ischaemia, heart diseases,
 CC atherosclerosis and endometriosis. The nucleic acid is also useful in
 CC drug screening, particularly for identifying agents for treating these
 CC pathologies
 XX XX
 XX Sequence 829 AA;

Query Match 99.8%; Score 4361; DB 6; Length 829;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 828; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGLPRGPLASLLQLQVWLQCAASPPCRAVFEAEVTLLEAGGAEQPGQALGKVFPMGCPG 60
 DB 1 MGLPRGPLASLLQLQVWLQCAASPPCRAVFEAEVTLLEAGGAEQPGQALGKVFPMGCPG 60
 QY 61 QEPALPSTNDNPTVRNGETVQERSLKERPLKIFPSKRILRRHKRWVAVIPSPENG 120
 DB 61 QEPALPSTNDNPTVRNGETVQERSLKERPLKIFPSKRILRRHKRWVAVIPSPENG 120
 QY 121 KGPPFQRLNQLKSKNDRTKIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDREEIAK 180
 DB 121 KGPPFQRLNQLKSKNDRTKIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDREEIAK 180
 QY 181 YELFHGAVSENGASVEDPMNISIIVTQNDHKPKFTQDTFRGSVLGVLPGTSVMQVAT 240
 DB 181 YELFHGAVSENGASVEDPMNISIIVTQNDHKPKFTQDTFRGSVLGVLPGTSVMQVAT 240
 QY 241 DEDDAIYTVNGVWVAYSIIHQSPKDPHDLMTFTHRSTGTSIVISSGLDREKPEYTLTQA 300
 DB 241 DEDDAIYTVNGVWVAYSIIHQSPKDPHDLMTFTHRSTGTSIVISSGLDREKPEYTLTQA 300
 QY 301 TDMCGGGTTTAVAVVEILDANDNAPMFDPOKYEAHVPENAVGHEVQRLTVDLDAPNSP 360
 DB 301 TDMCGGGTTTAVAVVEILDANDNAPMFDPOKYEAHVPENAVGHEVQRLTVDLDAPNSP 360
 QY 361 AWRATYLMGGDGDGHFTITTHPSNOGILTRKGLDPEAKNOHTLYVEVTNEAPVLUK 420
 DB 361 AWRATYLMGGDGDGHFTITTHPSNOGILTRKGLDPEAKNOHTLYVEVTNEAPVLUK 420
 QY 421 PTSTATIVHVEDVNEAPVFPVPSKVVEVQIGIPTGEPVVCVYTAEDDPKENOKISYRILR 480
 DB 421 PTSTATIVHVEDVNEAPVFPVPSKVVEVQIGIPTGEPVVCVYTAEDDPKENOKISYRILR 480
 QY 481 DPAGWLAMPDPSGQVAVGLDREDEQFVRNNIYEVNVLAMNDSPTTGTGILLTLID 540
 DB 481 DPAGWLAMPDPSGQVAVGLDREDEQFVRNNIYEVNVLAMNDSPTTGTGILLTLID 540
 QY 541 VNDHGPPPEPQITICNOSPVRHVNLITDKDLSPTSPFOAQLTDDSDIYVTAEVNEEGD 600
 DB 541 VNDHGPPPEPQITICNOSPVRHVNLITDKDLSPTSPFOAQLTDDSDIYVTAEVNEEGD 600
 QY 601 TVVLSLKKFLKQDVTYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFILLPV 660
 DB 601 TVVLSLKKFLKQDVTYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFILLPV 660
 QY 661 GAVLALLFLLLVLLVLRKKIKEPILLPEDDTRDNVYVYGGEGGEEDQDITQLHR 720
 DB 661 GAVLALLFLLLVLLVLRKKIKEPILLPEDDTRDNVYVYGGEGGEEDQDITQLHR 720

QY 721 GLEAPSEVLRNDVAPTIITPTMYRPRPNDPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
 Db 721 GLEAPSEVLRNDVAPTIITPTMYRPRPNDPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
 QY 781 DYEGSGDAASLSLTSSASDQDDYDYLNEWGSRFKKLADMYGGGDD 829
 Db 781 DYEGSGDAASLSLTSSASDQDDYDYLNEWGSRFKKLADMYGGGDD 829

RESULT 11
 ABUS6434
 ID ABUS6434 standard; protein; 829 AA.
 XX
 AC ABUS6434;
 XX
 DT 02-APR-2003 (first entry)
 DE Lung cancer-associated polypeptide #27.
 XX
 KW Lung cancer-associated polypeptide; cytostatic; emphysema;
 KW anti-inflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;
 KW small cell lung cancer; benign lesion; precancerous lesion; bronchitis;
 KW chronic obstructive pulmonary disease; hypersensitivity pneumonitis;
 KW interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.
 XX
 OS Unidentified.
 XX
 XX WO200286443-A2.
 PN
 XX
 PD 31-OCT-2002.
 XX
 XX 18-APR-2002; 2002WO-US012476.
 XX
 PR 18-APR-2001; 2001US-0284770P.
 PR 10-MAY-2001; 2001US-0250492P.
 PR 09-NOV-2001; 2001US-0339245P.
 PR 13-NOV-2001; 2001US-0350666P.
 PR 29-NOV-2001; 2001US-0334370P.
 PR 12-APR-2002; 2002US-0372246P.
 XX
 PA (EOSB-) EOS BIOTECHNOLOGY INC.
 XX
 PI Aziz N, Murray R;
 XX
 DR WPI; 2003-093161/08.
 DR N-PSDB; ABX76155.
 XX
 PT Detecting a lung cancer-associated transcript in a cell from a patient
 PT for treating lung cancer, by contacting a biological sample from the
 PT patient with a polynucleotide that exhibits increased or decreased
 PT expression in lung cancer.
 XX
 PS Claim 27; Page 210; 453pp; English.
 XX
 CC The invention relates to a method for detecting a lung cancer-associated
 CC transcript in a cell from a patient, comprising contacting a biological
 CC sample from the patient with a polynucleotide that selectively hybridizes
 CC to a sequence that is at least 80 % identical to a gene that exhibits
 CC increased or decreased expression in lung cancer samples. Lung cancer-
 CC associated polynucleotides and polypeptides are used for identifying a
 CC compound that modulates a lung cancer-associated polypeptide, for
 CC inhibiting proliferation of a lung cancer-associated cell to treat lung
 CC cancer in a patient and for treating a mammal having lung cancer by
 CC administering a modulatory compound identified. The methods are useful
 CC for treating lung cancer, such as small cell lung cancer, non-small cell
 CC lung cancer or other benign or precancerous lesions, e.g. atelectasis,
 CC emphysema, bronchitis, chronic obstructive pulmonary disease, fibrosis,
 CC hypersensitivity pneumonitis, interstitial pulmonary fibrosis, asthma and
 CC bronchiectasis. The genes, polynucleotides and polypeptides are useful
 CC for diagnostic purposes and as targets for screening for therapeutic
 CC compounds that modulate lung cancer, such as antibodies. Sequences
 CC ABUS6408-ABUS6745 represent lung cancer-associated polypeptides of the
 CC invention

XX Sequence 829 AA;
 SQ
 Query Match 99.8%; Score 4361; DB 6; Length 829;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 828; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MGLPRGPLASLLLLQVCWLQCAASEPCRAVFEAEVLEAGGAEQEGQALGVKVMGCPG 60
 Db 1 MGLPRGPLASLLLLQVCWLQCAASEPCRAVFEAEVLEAGGAEQEGQALGVKVMGCPG 60
 QY 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKILRRHKDDWVAPISVPENG 120
 Db 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKILRRHKDDWVAPISVPENG 120
 QY 121 KGPFPORLNQKSNKDRDTKIFYITGPADSPPEGVFAVEKETGWLKLNKPLDREIAK 180
 Db 121 KGPFPORLNQKSNKDRDTKIFYITGPADSPPEGVFAVEKETGWLKLNKPLDREIAK 180
 QY 181 YELFGHAYSENGASVEDPMNISIIIVTDQNDHKPFTQDTFRGSVLGVLPGTSMQVAT 240
 Db 181 YELFGHAYSENGASVEDPMNISIIIVTDQNDHKPFTQDTFRGSVLGVLPGTSMQVAT 240
 QY 241 DEDDAIYTYNGVAYSTHSQSPKDPHDLMTFTHSTGTISVISSGLDREKVPYTLTQIA 300
 Db 241 DEDDAIYTYNGVAYSTHSQSPKDPHDLMTFTHSTGTISVISSGLDREKVPYTLTQIA 300
 QY 301 TDMGSGSTTTTAVAVVEILDANDNAMPDPKPYEAHVPENAVHVEVORLTVTDLDA NSP 360
 Db 301 TDMGSGSTTTTAVAVVEILDANDNAMPDPKPYEAHVPENAVHVEVORLTVTDLDA NSP 360
 QY 361 AWRATYLLMGDDGDGHFTITTHPSNQILITRGLDPEAKNOHTLYVEVTNEAPFVKL 420
 Db 361 AWRATYLLMGDDGDGHFTITTHPSNQILITRGLDPEAKNOHTLYVEVTNEAPFVKL 420
 QY 421 PTSTATIVVHVVEDVNEAPVFPSPKVEVEQGIPTGPPVCVYTAEDPKENQKISYILR 480
 Db 421 PTSTATIVVHVVEDVNEAPVFPSPKVEVEQGIPTGPPVCVYTAEDPKENQKISYILR 480
 QY 481 DPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVWVLAMDNQSPPTTGTLLTLLID 540
 Db 481 DPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVWVLAMDNQSPPTTGTLLTLLID 540
 QY 541 VNDHGPVPEPRQITTCNOSPVRHVLNITDKLSPTSPFOAQLTDDSDIYWTAEVNEEGD 600
 Db 541 VNDHGPVPEPRQITTCNOSPVRHVLNITDKLSPTSPFOAQLTDDSDIYWTAEVNEEGD 600
 QY 601 TVVLSLKKFLKQDVTYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGFILPVL 660
 Db 601 TVVLSLKKFLKQDVTYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGFILPVL 660
 QY 661 GAVTALLFLLVLLVLLVLRKSKIKPEPLLPEDDTRDNVYFYGGEGGEDDQDITQLHR 720
 Db 661 GAVTALLFLLVLLVLLVLRKSKIKPEPLLPEDDTRDNVYFYGGEGGEDDQDITQLHR 720
 QY 721 GLEAPSEVLRNDVAPTIITPTMYRPRPNDPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
 Db 721 GLEAPSEVLRNDVAPTIITPTMYRPRPNDPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
 QY 781 DYEGSGDAASLSLTSSASDQDDYDYLNEWGSRFKKLADMYGGGDD 829
 Db 781 DYEGSGDAASLSLTSSASDQDDYDYLNEWGSRFKKLADMYGGGDD 829

RESULT 12
 ABUS6670
 ID ABUS6670 standard; protein; 829 AA.
 XX
 AC ABUS6670;
 XX
 DT 02-APR-2003 (first entry)
 XX Lung cancer-associated polypeptide #263.
 DE

KW	Lung cancer-associated polypeptide; cytostatic; emphysema;	QY	181	YELFHGAVSENGASVEDPMNIIIVTDQNDHAKPKTQDTFRGSVLEGLVPGTSMQVMTAT	240
KW	anti-inflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;	Db	181	YELFHGAVSENGASVEDPMNIIIVTDQNDHAKPKTQDTFRGSVLEGLVPGTSMQVMTAT	240
KW	small cell lung cancer; benign lesion; precancerous lesion; bronchitis;	QY	241	DEDDAIYVNGVAVYSIHSQEPKPDHMLFTTHRSTGTISVSSGLDREKVPYETLTIOA	300
KW	chronic obstructive pulmonary disease; hypersensitivity pneumonitis;	Db	241	DEDDAIYVNGVAVYSIHSQEPKPDHMLFTTHRSTGTISVSSGLDREKVPYETLTIOA	300
KW	interstitial pulmonary fibrosis; fibrosis; aschma; bronchiectasis.	QY	301	TDMDGSGSTTTAVAVVEILLDANDNAPMDPQKYEAHVPENAVGHVQRLTVTDLDPNSP	360
XX	Unidentified.	Db	301	TDMDGSGSTTTAVAVVEILLDANDNAPMDPQKYEAHVPENAVGHVQRLTVTDLDPNSP	360
XX	WO200286443-A2.	QY	361	AWRATYLMGDDGDHFTTTHPESNQGLTTRKGLDPEAKNQHTLYVEVNEAPFVLKL	420
XX	31-OCT-2002.	Db	361	AWRATYLMGDDGDHFTTTHPESNQGLTTRKGLDPEAKNQHTLYVEVNEAPFVLKL	420
XX	18-APR-2002; 2002WO-US012476.	QY	421	PTSTATIVVHVEDVNEAPVFPVPSKVVEQEGIPTEGVCVVTAEDDPKENOKISYRIIL	480
XX	18-APR-2001; 2001US-0284770P.	Db	421	PTSTATIVVHVEDVNEAPVFPVPSKVVEQEGIPTEGVCVVTAEDDPKENOKISYRIIL	480
XX	10-MAY-2001; 2001US-0290492P.	QY	481	DPAGMLAMPDPSGQVTAAGTLDREDEQFVRNNIYEVWVLMNDNGSPPTTGTGLLLTLID	540
XX	09-NOV-2001; 2001US-0339245P.	Db	481	DPAGMLAMPDPSGQVTAAGTLDREDEQFVRNNIYEVWVLMNDNGSPPTTGTGLLLTLID	540
XX	13-NOV-2001; 2001US-0350666P.	QY	541	VNDHGFVPEPROITICNQSPVRHVLNITDKOLSPHTSPFOAQLTDDSDIYWTAEVNEEGD	600
XX	29-NOV-2001; 2001US-0334370P.	Db	541	VNDHGFVPEPROITICNQSPVRHVLNITDKOLSPHTSPFOAQLTDDSDIYWTAEVNEEGD	600
XX	12-APR-2002; 2002US-0372246P.	QY	601	TVVLSLKFLKQDTYDVHLSLDHGNKQLTIVIRATVCDCHGVETCPGPKWGFIPLV	660
XX	(EOSB-) EOS BIOTECHNOLOGY INC.	Db	601	TVVLSLKFLKQDTYDVHLSLDHGNKQLTIVIRATVCDCHGVETCPGPKWGFIPLV	660
XX	Aziz N, Murray R;	QY	661	GAVALALLFLLVLLVLRKKRKKIPEPLLPEDDTDRDNVFFYEGEGGGBEDQDYDTQLHR	720
XX	WPI; 2003-093161/08.	Db	661	GAVALALLFLLVLLVLRKKRKKIPEPLLPEDDTDRDNVFFYEGEGGGBEDQDYDTQLHR	720
XX	N-PSDB; ABX76399.	QY	721	GLEARPEVLRNDVAPTIIPTWYPRPANDEIGNFIENLKAANTDPTAPPYDTLLVF	780
XX	Detecting a lung cancer-associated transcript in a cell from a patient	Db	721	GLEARPEVLRNDVAPTIIPTWYPRPANDEIGNFIENLKAANTDPTAPPYDTLLVF	780
XX	for treating lung cancer, by contacting a biological sample from the	QY	781	DYEGSGSDAASLSLTSSASDQDQDYDLNENWGSFRFKKLADMYGGGEDD	829
XX	patient with a polynucleotide that exhibits increased or decreased	Db	781	DYEGSGSDAASLSLTSSASDQDQDYDLNENWGSFRFKKLADMYGGGEDD	829
XX	expression in lung cancer.	XX	RESULT 13		
XX	Claim 27; Page 392; 453pp; English.	XX	AY09375		
XX	The invention relates to a method for detecting a lung cancer-associated	XX	ID	AY09375 standard; protein; 882 AA.	
XX	transcript in a cell from a patient, comprising contacting a biological	XX	XX	AY09375;	
XX	sample from the patient with a polynucleotide that selectively hybridises	XX	AC	AY09375;	
XX	to a sequence that is at least 80 % identical to a gene that exhibits	XX	DT	14-JUL-1999 (first entry)	
XX	increased or decreased expression in lung cancer samples. Lung cancer-	XX	DE	Wild-type E-cadherin.	
XX	associated polynucleotides and polypeptides are used for identifying a	XX	XX	E-cadherin; Maori; familial gastric cancer; germline mutation; detection;	
XX	compound that modulates a lung cancer-associated polypeptide, for	XX	KW	human; breast cancer; colorectal cancer; prostate cancer; thyroid cancer;	
XX	inhibiting proliferation of a lung cancer-associated cell to treat lung	XX	KW	kidney cancer; bladder cancer; liver cancer;	
XX	cancer in a patient and for treating a mammal having lung cancer by	XX	KW	hereditary diffuse gastric cancer; HDGC.	
XX	administering a modulatory compound identified. The methods are useful	XX	XX	Homo sapiens.	
XX	for treating lung cancer, such as small cell lung cancer, non-small cell	XX	OS	WO9920168-A2.	
XX	lung cancer or other benign or precancerous lesions, e.g. atelectasis,	XX	PN	29-APR-1999.	
XX	emphysema, bronchitis, chronic obstructive pulmonary disease, fibrosis,	XX	XX	19-OCT-1998; 98WO-NZ000160.	
XX	hypersensitivity pneumonitis, interstitial pulmonary fibrosis, asthma and	XX	XX	17-OCT-1997; 97NZ-00328994.	
XX	bronchiectasis. The genes, polynucleotides and polypeptides are useful	XX	XX	(UYOT-) UNIV OTAGO.	
XX	for diagnostic purposes and as targets for screening for therapeutic	XX	PA	(TEWH-) TE WHETU WHANAU TRUST LTD.	
XX	compounds that modulate lung cancer, such as antibodies. Sequences				
XX	ABU56408-ABU56745 represent lung cancer-associated polypeptides of the				
XX	invention				
SQ	Sequence 829 AA;				
Query Match	99.6%; Score 4353; DB 6; Length 829;				
Best Local Similarity	99.8%; Pred. No. 0;				
Matches 827; Conservative	0; Mismatches 2; Indels 0; Gaps 0;				
Qy	1 MGLPRGLASLLLLQVCLWQCAASEPCRAVFREAEVTLAAGAEQEPGQALGKVFVPGCPG	60			
Db	1 MGLPRGLASLLLLQVCLWQCAASEPCRAVFREAEVTLAAGAEQEPGQALGKVFVPGCPG	60			
Qy	61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIPFSKILRRHKDWTVAISVPENG	120			
Db	61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIPFSKILRRHKDWTVAISVPENG	120			
Qy	121 KGPPFPQRLNOLKSNKRDTKIFYSITGPGADSPPEGVFAVEKETGWLILNKLDPDREIAK	180			
Db	121 KGPPFPQRLNOLKSNKRDTKIFYSITGPGADSPPEGVFAVEKETGWLILNKLDPDREIAK	180			

XX PI Reeve AE, Guilford PJ;
XX DR WPI; 1999-288129/24.
XX DR N-PSDB; AAX56022.
XX PT Determining predisposition to cancer by detecting a mutation in the E-cadherin gene.
XX PT Disclosure; Page 35-37; 55pp; English.
XX FS The present invention a method for detecting the presence or absence of mutations in the E-cadherin gene, which is indicative of a predisposition to cancer. The method can be used to identify predisposition to cancers such as breast cancer, colorectal cancer, gastric cancer, prostate cancer, thyroid cancer, kidney cancer, bladder cancer, and liver cancer. The method is particularly useful for identifying predisposition to hereditary diffuse gastric cancer (HDGC). Compounds which increase the expression or prevent the decrease of E-cadherin would be potential cancer chemopreventive agents. Gene therapy can also be used to supply wild-type E-cadherin. The key to cancer treatment is early detection. The method allows the identification of individuals with a predisposition to cancer, particularly to hereditary cancer, which enables detection before the appearance of clinical symptoms, and thus allows treatment, or other courses of action, to commence as soon as possible. Also, families with histories of familial cancer will be able to undergo tests to search for E-cadherin gene mutations. The present sequence represents wild-type E-cadherin
XX CC
XX SQ Sequence 882 AA;

Query Match 55.4%; Score 2420; DB 2; Length 882;
Best Local Similarity 55.1%; Pred. No. 2.2e-192;
Matches 483; Conservative 112; Mismatches 222; Indels 60; Gaps 13;

QY 9 ASLLLLQV-CWLOCAAEPCRAVREABVLEAGGAQEGQALGKV-FMGCPG-QEPAL 65
DB 10 ALLLLLQVSSWL-CQEPCHPGFDAESYFTVPRHLERGRVLGRVNFEDCTGRQTAY 68
QY 66 FSTDNDDFTVANGETVOERRSLKERNP----- 92
DB 69 FSLDT-RFKVGTGDVIVTKRPLRHNQIHLVYANDSTYKSTKYLNTVGHHRPP 127
QY 93 -----LKTFP-SKRLRHKDWWVAPISVPENKGGPPQRLNOLKNKORDTK 140
DB 128 HQASVSGIQABELTFPNSPGLRQKRDWIPPIPCPENKGGPPKVLQVQIKNDKEG 187
QY 141 IFYSTIGPADSPPEGFAVEKETGWLNNKPLDREIAKVELFGHAVSENGASVEDPMN 200
DB 188 VFYSITGGADTPPVGVFIIBRETKWLVTEPLDRERLAIYTLFSAVSNNGNAVEDPME 247
QY 201 ISITVQNDHKPFTQDTERGSLVLEGLPQTSVMQVATDEDDAIYTYNGVAYSHSQ 260
DB 248 ILITVTDQNDKPEFTQEVFGKSVMEGALPQTSVMEVATDADDVNTYNAAIYTLISQ 307
QY 261 EPKPDHLMFTHIRSTGTISVISGLDREKVPETLTIQATDMDGDSITTAVAVVILD 320
DB 308 DPELPDKMFTINRNTGVISVVTTLGLDRESPTTLVQADLOGEGLSTIATAVITVD 367
QY 321 ANDNAPMDPKYBAHPENAVGHEVORLVTDLDPNSPAWRATYILIMGGDGDHFTIT 380
DB 368 TNDNPPINPTTKGVQVPEANVVIITLKVTDADAPNTFAWEAVYTL-NDGGQFVVT 426
QY 381 THPSNQILITRGLDFEAKNQHTLVVETNEAPFVLKPTSTATVWVEDWNEAPVF 440
DB 427 TNPVNDGILKTAKGLDFEAKQQYILHVAVTNVVPFVSUFTTSTATVTDVLDVNEAPIF 486
QY 441 VPPSKVVEQGIPTGPEVCVYTAEDDDK-ENQKISYRILRDPAGWLAMPDSDGAVG 499
DB 487 VPPEKRVESDFGVGEITSYTAQEDPTFMEQKITVIRWDTANWLEINFDTGAISTRA 546
QY 500 TLREDEGQFVRNNIYVWVLMVNDGNSPPTTGTGTLILLTLIDVNDHGVDPSPRQITICNQ 559
DB 500 TLREDEGQFVRNNIYVWVLMVNDGNSPPTTGTGTLILLTLIDVNDHGVDPSPRQITICNQ 559

DB 547 ELDRDPFHVKNSVTYALIIATDNGSPVATGTGTLILLSDVNDNAPIPEPTIFFCERN 606
QY 560 PVRHVLNITKDLSPHTSPFOAQLTDDSDIYVTAEVNE-EGDTVVLSSKKFLKQDYDVH 618
DB 607 PKPQVINIADLDPNTSPFTAEILTHGASANWTQYNDPTQESIIILPKMALEVGDYKIN 666
QY 619 LSLSDHGNKEQLTVIRATVCDCHGVETC--PGPWKGGF----LPVLGAVIALLLFLLV 672
DB 667 LKLMNQNKQDQVTLVSVCDCEGAAGVCRKAQVCEAGLQIPAILGILGILALLILL 726
QY 673 LLLLVRKRNKKEPLLPEDDTRDNVFFYEGGEGEDQDYITQLHRLGAREVVRN 732
DB 727 LLLFLRRRAVVKPELLPPEDDTRDNVFFYEGGEGEDQDFLSQLHRLGLDAREVTRN 785
QY 733 DVAPITIIPTMYRPRPANDPDEIGNFIENLKAANTDTPAPPYDILLVDFYEGSGSAA 792
DB 786 DVAPITMSVPRYLPANPDEIGNFIENLKAANTDTPAPPYDILLVDFYEGSGSAA 845
QY 793 SSLTSASDQDQDYDYLNEWSRFFKCLADMYGGGEDD 829
DB 846 SSLNSESXKQDYDYLNEWSRFFKCLADMYGGGEDD 882

RESULT 14
AAB35730
ID AAB35730 standard; protein; 882 AA.
XX AC AAB35730;
XX DT 19-FEB-2001 (first entry)
XX DE Human E-cadherin amino acid sequence.
XX KW Pharmacological composition; gene therapy; junctional complex disruption; cytostatic; cardiant; vasotrophic; antiinflammatory; antitipaeamic; cystic fibrosis; inflammatory disease; fibrotic lung disease; peripheral vascular disease; coronary arterial disease; restenosis; hypercholesterolaemia; cancer; RGD; human; E-cadherin.
XX OS Homo sapiens.
XX PN WO200062815-A2.
XX PD 26-OCT-2000.
XX PF 13-APR-2000; 2000WO-GB001408.
XX PR 15-APR-1999; 99GB-00008636.
PR 15-APR-1999; 99GB-00008643.
PR 20-APR-1999; 99GB-00009073.
PR 22-APR-1999; 99GB-00009297.
PR 13-SEP-1999; 99US-0153757P.
XX (GLAX) GLAXO GROUP LTD.
PA (UTNC-) UNIV NORTH CAROLINA.
XX Boucher RC, Ford MJ, Johnson LG, Man Y, West MR;
PI WPI; 2000-679557/66.
XX DR N-PSDB; AAC66247.
XX PT Novel pharmaceutical composition useful for gene therapy, comprises nucleic acid encoding therapeutic gene and a promoter, transfection vehicle and an agent which disrupts function of junctional complex in cells.
XX PS Disclosure; Fig A; 51pp; English.
XX CC This invention relates to a pharmaceutical composition used for gene therapy, comprising a nucleic acid encoding a therapeutic gene and a promoter, a transfection vehicle and an agent to disrupt the function of the junctional complex in the cells. The pharmaceutical composition has cytostatic, cardiant, vasotrophic, antiinflammatory and antitipaeamic

Sequence 882 AA;

Query Match 55.4%; Score 2420; DB 4; Length 882;
Best Local Similarity 55.1%; Pred. No. 2.2e-192;
Matches 483; Conservative 112; Mismatches 222; Indels 60; Gaps 13;

QY 9 ASLLQLQV-CHLQCAASEPCRAVREAEVTLLEAGAEQEPQALGKV-FMGCPG-QEPAL 65
DB 10 ALLLLQLVSSL-COEPEPCHPGDAESYFTVPRHLERGLRVGNFEDCTGORTAY 68
QY 66 FSTNDNFTVANGETVQERRSLKERNP----- 92
DB 69 FSLDT-RFKVGTGVIITVKPLRPHNQIHFLVYAWDSTYRKSTKVTLTNTVGHHRPPP 127
QY 93 -----LKTFP-SKRILRHKRDWVAPISVDPENKGPFPQRLNOLKSKORDTK 140
DB 128 HQASVSGIAELLTFPNSSPGLRRQKRDWIPPIPCPENKGPFPKVLQVQIKSKDKKEG 187
QY 141 IFYSITGPGADSPPEGVFAVEKETGMLLNKPLDREBIKVELFHHAVSNGASVEDPMN 200
DB 188 VFSITGGGADTPVGVIIFERIEICWLVTEPLDREIATYTLFHAVSSNGNAVEDPME 247
QY 201 ISIIVTQNDHKPKFTQDTFRGSLVGLVPGTSVMQVTATDEDDAIYTYNGVVAYSIHSQ 260
DB 248 ILITVTDQNDKPEFTQEVFKGSMEGALPGTSVMEVTATDADDVNTYNAIAYTILSQ 307
QY 261 EPKDPHDMFTIHRSTGTISVISSGLDREKVEYTLTIQATMDGDGSTTTAVAVVEILD 320
DB 308 DPELPDKMFTINRTGIVSVITGLDRESFPYTLVQAADLQGEGLSTTATAVITVD 367
QY 321 ANDNAPFDPQKYEAHVPENAVGHEVQRLTVDLDAENSPAWRATYLLIMGGDDGDHFTIT 380
DB 368 TNDNPPFNPFTYKQVPEANEANVITTLKVTADAPNTPAWEAVTIL-NDGGQFVVT 426
QY 381 THPSNQGILTRGLDPEAKNOHTLVVEVTNEAPFVLKLPSTATIVHVEDVNEAPVF 440
DB 427 TNPVNDGILKTAGLDPEAKQYILHVAVTNVVPEVSITSTATVTVDLVDNEAPIF 486
QY 441 VPPSKWVEVQGIPTGSPVCVYTAEDPK-ENKISYRIILDPAGWLAMPDSQVTVAG 499
DB 487 VPPEKRVESDFVGQEIISYTAQEPDTEMEQKITVIRWDTANWLEINPDGAISTRA 546
QY 500 TLDREDEQFVRNNIYEVWVLAAMDNGSPPTGTGTLTLLTLIDVNDHGPVPEPROITCNQS 559
DB 547 ELDRDEFHVKNSTYATALLIATDNGSPVATGTGTLTLLILSDVNDNAPIPEPRTIFFCERN 606
QY 560 PVRHVLNITDKLSPHTSPFOAQLTDDSDIYWTAEVNE-EGDVTVLSLKKFLKQDTYDVH 618
DB 607 PKQVINIIDLPLPNTSPFPAELTHGASANWTIOYNDPTQESIIILKPKMALEVGDKIN 666
QY 619 LSLSDHGNKEQLTVIRATVCDCHGVETC--PGPWKGF----ILPVLGAVLALLFLLV 672
DB 667 LKMDNQNKDQVTTLEVSVCDCGAAGVCRKAQVVEAGLQIPAILGILGGILALLIILL 726
QY 673 LLLLVKKEKIKPELLPEDDTRNVFYEGEGGEDQDYDITOLHRLGARPEVVLN 732
DB 727 LLLFLRRVAVKPELLPEDDTRNVYIYDEEGGEDQDFDLSQLHRLGARPEVT-RN 785
QY 733 DVAPTIPTMYRPRANPDEIGNFIENLKAANTDPTAPPYDTLLVFDYEGSGSAAASL 792
DB 786 DVAPTLMSVPRYLPRANPDEIGNFIDENLKAADTPTAPPYDSSLVFDYEGSGSEASL 845
QY 793 SSLTSSASDQDDYDYLNEGSRFKKLADMYGGEDD 829
DB 846 SSLNSSESDDQDYDYLNEGWRFKKLADMYGGEDD 882

OM protein - protein search, using sw model

Run on: September 21, 2004, 22:04:29 ; Search time 21 Seconds
(without alignments)

(without alignments)
2037,997 Million cell updates/sec

J5-09-916-849A-1

Perfect score:

Sequence: 1 MGLPRGPLASLLLLQVCWLQ.....NEWGSRFKKLADMYGGGEDD 829

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

```

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

```

Post-processed: Minimum Match 0%

Post-processing: Minimum Match 0%
Maximum Match 100%

Maximum Match 100%
Testing first 45 summaries

Database : Issued Patents 22.*

```
Database : ISSUED_PATENTS_AA:
1. /com2 6/ptdata/2/iaa/5A COMB ren.*
```

```

1: /cgn2_b/pdata/2/1aa/3A_COMB.pdb:
2: /cgn2_b/pdata/2/1aa/5B_COMB.pdb:

```

```
2: /cgn2_6/protdata/2/1aa/3b__COMB.per:*
3: /cgn2_6/protdata/2/1aa/6a__COMB.per:*
```

```
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pdb: *
4: /cgn2_6/ptodata/2/1aa/6H_COMB.pdb: *
```

```
4: /cgn2_6/ptodata/2/1aa/6B_COMB.per:*
5: /cgn2_6/ptodata/2/1aa/6B_COMB.per:
```

```
5: /cgn2_6/ptodata/2/1aa/PTUS_COMB.dat
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			Length	DB	ID	Description
	Score	Match					
1	3553.5	81.3	822	2	US-08-474-067-9		Sequence 9, Appli
2	3553.5	81.3	822	2	US-08-474-068A-9		Sequence 9, Appli
3	3553.5	81.3	822	2	US-08-474-481-8		Sequence 8, Appli
4	2384	54.6	884	2	US-08-474-067-8		Sequence 8, Appli
5	2384	54.6	884	2	US-08-474-068A-8		Sequence 8, Appli
6	2384	54.6	884	2	US-08-472-481-7		Sequence 7, Appli
7	2348.5	53.8	878	1	US-08-237-919-2		Sequence 2, Appli
8	2348.5	53.8	878	4	US-08-732-423-2		Sequence 2, Appli
9	2348.5	53.8	878	4	US-09-798-267-2		Sequence 2, Appli
10	2348.5	53.8	878	4	US-09-798-267-3		Sequence 3, Appli
11	2348.5	53.8	878	5	PCF-US95-05518-2		Sequence 2, Appli
12	2231.5	51.1	837	2	US-08-474-067-7		Sequence 7, Appli
13	2231.5	51.1	837	2	US-08-474-068A-7		Sequence 7, Appli
14	2231.5	51.1	837	2	US-08-472-481-6		Sequence 6, Appli
15	1714	39.2	906	4	US-09-417-039-11		Sequence 11, Appl
16	1694	38.8	913	2	US-08-474-067-6		Sequence 6, Appli
17	1694	38.8	913	2	US-08-474-068A-6		Sequence 6, Appli
18	1694	38.8	913	2	US-08-472-481-5		Sequence 5, Appli
19	1621.5	37.1	916	1	US-08-188-228-48		Sequence 48, Appl
20	1621.5	37.1	916	1	US-08-332-643-42		Sequence 42, Appl
21	1621.5	37.1	916	1	US-08-332-638-48		Sequence 48, Appl
22	1175.5	26.9	555	2	US-08-453-702A-98		Sequence 98, Appl
23	1164	26.6	556	1	US-07-998-003A-98		Sequence 98, Appl
24	1164	26.6	556	1	US-08-453-274B-98		Sequence 98, Appl
25	1164	26.6	556	1	US-08-453-695A-98		Sequence 98, Appl
26	1164	26.6	556	1	US-08-268-161A-98		Sequence 98, Appl
27	1164	26.6	556	3	US-09-099-633-98		Sequence 98, Appl

Query Match	81.3%;	Score 3553.5;	DB 2;	Length 822;
Best Local Similarity	81.5%;	Pred. No. 0;		
Matches 678;	Conservative	58; Mismatches	83; Indels	13; Gaps

1 MGLPRGPLA-SLLLLQVCWLQCAASEPCRAVF-REAETLEAGGAEQEPGQALGKVFMC 58

28	1023.5	23.4	713	1	US-08-188-228-62	Sequence 52, Appl
29	1023.5	23.4	713	1	US-08-332-643-56	Sequence 567, Appl
30	1023.5	23.4	713	1	US-08-332-638-62	Sequence 62, Appl
31	976	22.3	712	2	US-08-474-067-2	Sequence 2, Appl
32	976	22.3	712	2	US-08-474-067-5	Sequence 5, Appl
33	976	22.3	712	2	US-08-474-067-5	Sequence 5, Appl
34	976	22.3	712	2	US-08-474-068A-2	Sequence 5, Appl
35	976	22.3	712	2	US-08-474-068A-5	Sequence 5, Appl
36	976	22.3	712	2	US-08-472-481-2	Sequence 2, Appl
37	976	22.3	717	2	US-08-474-067-4	Sequence 4, Appl
38	976	22.3	717	2	US-08-474-068A-4	Sequence 4, Appl
39	968.5	22.2	776	2	US-08-472-483-4	Sequence 4, Appl
40	966.5	22.1	796	1	US-08-738-349-4	Sequence 4, Appl
41	966.5	22.1	796	1	US-08-188-228-58	Sequence 58, Appl
42	966.5	22.1	796	1	US-08-332-643-52	Sequence 52, Appl
43	951.5	21.8	796	2	US-08-332-638-58	Sequence 58, Appl
44	933	21.4	799	1	US-08-738-349-2	Sequence 2, Appl
45	933	21.4	799	1	US-08-188-228-42	Sequence 42, Appl
46	933	21.4	799	1	US-08-332-638-42	Sequence 42, Appl

ALIGNMENTS

RESULT 1
US-08-474-067-9
Sequence 9, Application US/08474067
Patent No. 5811518
GENERAL INFORMATION:
APPLICANT: Ranscht, Barbara
TITLE OF INVENTION: T-Cadherin Adhesion Molecule
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESS: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
City: San Diego
STATE: California
COUNTRY: United States
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,067
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/213,361
FILING DATE: 14-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/607,293
FILING DATE: 30-OCT-1990
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 1682
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 822 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-474-067-9

Query Match 81.3%: score 3553.5; DB 2; Length 822;

QY 658 PVLGAVLALLFLLVLLLVKRRKIKKEPLLPEDDTRDNVYFGEGBEDQDYDITQ 717
DB 651 PIUGAVLALLFLLVLLLVKRRKIKKEPLLPEDDTRDNVYFGEGBEDQDYDITQ 710
QY 718 LHRGLEARPEVLRNDVAPTIPTMYRPRPANPDEIGNFIENLKAANTDTPAPPYDL 777
DB 711 LHRGLEARPEVLRNDVAPTIPTMYRPRPANPDEIGNFIENLKAANTDTPAPPYDL 770
QY 778 LVFDYEGSGSDAASLSLTSASDQDQDYDYLNEWGSRKKLADMYGGGEDD 829
DB 771 MVFDYEGSGSDAASLSLTSASDQDQDYDYLNEWGSRKKLADMYGGGEDD 822

RESULT 3
US-08-472-481-8
; Sequence 8, Application US/08472481
; Patent No. 5863804
; GENERAL INFORMATION:
; APPLICANT: Ranscht, Barbara
; TITLE OF INVENTION: T-Cadherin Adhesion Molecule
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,481
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/213,361
; FILING DATE: 14-MAY-1994
; PRIOR APPLICATION NUMBER: US 07/607,293
; FILING DATE: 30-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1686
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 822 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-472-481-8

Query Match 81.3%; Score 3553.5; DB 2; Length 822;
Best Local Similarity 81.5%; Pred. No. 0;
Matches 678; Conservative 58; Mismatches 83; Indels 13; Gaps 5;

QY 1 MGLPRGPLA-SLLLLQVWLQCAASEPCRAVF-REARVTLAAGAGQEPQOALGKVFMC 58
DB 1 MELLUGPHAFILLLLQVWLRSVSEPRAGFVGEAGVTLVETGDLPSQVLGKVALAG 60
QY 59 PGQBPALFSTDNDFTVRNGETVQERSLKERNPDKIPPSKRIILRRHKRWVVAISVPE 118
DB 61 QGMEHA---DNGDIIMLTGRVQGGKAMHS-----PPTRIILRRKRWVMPPIFVPE 110
QY 119 NGKGPFFORLNQLKSNDRGKTKFYSITGCAADPPGKFAVEKCTGMLLNKPLDREET 178
DB 111 NGKGPFFORLNQLKSNDRGKTKFYSITGCAADPPGKFAVEKCTGMLLNKPLDREKI 170

QY 179 AKYELFGHAVSENGASVEDPMNIIIVTDQNDHKPKFTQDTRFGSVLEGVLPGTSMQVT 238
DB 171 VKYELGHAVSENGASVEEPMNIIIVTDQNDNPKFTQDTRFGSVLEGVLPGTSMQVT 230
QY 239 ATDEDDALYTVNGVAVSIHSEQEPKPDHLMFTTHRSTGTSTISVSSGLDRKKVPEYILT 298
DB 231 ATDEDDAVNTYNGVAVSIHSEQEPKPDHLMFTTHKSTGTSTISVSSGLDRKKVPEYILT 290
QY 299 QATMDGSGSTTAVVVEILDANDNAPMFPQKYEAHVPENAVGVHEVQRVLTVDLDAVN 358
DB 291 QATMDGSGSTTAEAVVQIILDANDNAPFEPQKYEAHVPENAVGVHEVQRVLTVDLDAVN 350
QY 359 SPAMRATYLMGGDDGCHFTTTHPEQNQGLITRKLGLDFEAKNQHTLYVEVNTNEAPVL 418
DB 351 WPAWRATYHVGDDGCHFTTTHPETNQGLITRKLGLDFEAKNQHTLYVEVNTNEAPVL 410
QY 419 KLPTSTATIVVHVDNEAPVFPSPKIVEQEGIPTEGPEVCVTAEDPDKENOKISYRI 478
DB 411 KLPTATATVTVVHVVDNEAPVFPSPKIVEQEGISIGELVCIYTAQDPDKEDQKISYTI 470
QY 479 LRDPAGWLAMPDPSGQVTAAGTLDREDEQFVKNVYEVWLAMDNGSPPTTGTGTLTLL 538
DB 471 SRDPANWLAVDPDSQITAAAGILDREDEQFVKNVYEVWLAMDNGSPPTTGTGTLTLL 530
QY 539 IDVNDHGPVPEPRQITICNQPVRHVLNITDKDLSPHSTPFOAQLTDDSDIYNTAEVNEE 598
DB 531 TDINDHGPPEPRQIIICNQPVPQVLNITDKDLSPNSSPFOAQLTDDSDIYNTAEVNEE 590
QY 599 GDTVVLKXFLKQDITYVHLSLSDHGNKEQLTWRATVCDCHGHV-ETCPGPKWGGFLL 657
DB 591 GDTVALSLKXFLKQDITYDLHLSLSDHGNKEQLTWRATVCDCHGHVNDPCKPRWKGFFIL 650
QY 658 PVLGAVLALLFLLVLLLVKRRKIKKEPLLPEDDTRDNVYFGEGBEDQDYDITQ 717
DB 651 PIUGAVLALLFLLVLLLVKRRKIKKEPLLPEDDTRDNVYFGEGBEDQDYDITQ 710
QY 718 LHRGLEARPEVLRNDVAPTIPTMYRPRPANPDEIGNFIENLKAANTDTPAPPYDL 777
DB 711 LHRGLEARPEVLRNDVAPTIPTMYRPRPANPDEIGNFIENLKAANTDTPAPPYDL 770
QY 778 LVFDYEGSGSDAASLSLTSASDQDQDYDYLNEWGSRKKLADMYGGGEDD 829
DB 771 MVFDYEGSGSDAASLSLTSASDQDQDYDYLNEWGSRKKLADMYGGGEDD 822

RESULT 4
US-08-474-067-8
; Sequence 8, Application US/08474067
; Patent No. 5811518
; GENERAL INFORMATION:
; APPLICANT: Ranscht, Barbara
; TITLE OF INVENTION: T-Cadherin Adhesion Molecule
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,067
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/213,361
; FILING DATE: 14-MAY-1994

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/607,293
; FILING DATE: 30-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1682
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 884 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-474-067-8

Query Match          54.6%; Score 2384; DB 2; Length 884;
Best Local Similarity 53.5%; Pred. No. 7e-213;
Matches 471; Conservative 123; Mismatches 221; Indels 66; Gaps 13;

QY 9 ASLLLLQV-CWLQCAASEP--CRAVFREAETLEAGAEQEPGQALGKV-FMGCPCQGEPA 64
DB 10 ALLLLQVSSWL-COELEPESCSGFSSEVYFPVPERHLERGHVLRGVFEGCTGRPT 68
QY 65 LFSTDNDFTVRNGETVQERRSLK----- 88
DB 69 AFFSEDSRFKATDGTITVKEHLKHLKLETSFLVRARDSSHRELSTKVTLKSMGHHHRH 128
QY 89 -----ERNP-LKIFPSKRI-LRRHKRDVWVAPISVPEKNGKGFPPORLNOLKSNKDRDT 139
DB 129 HHRDPAGESNELLMFFSVPLGRQRKRDWVIPPISCPENEKGEFPKNLVQIKSNDRKET 188
QY 140 KIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDREETIAKYELFQHAVSNGASVEDPM 199
DB 189 KVFYSITGQADKPEVGVGFIERETGLWKTQPLDREAIKAYILYSHAVSNGEVEDPM 248

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/607,293
; FILING DATE: 30-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1683
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 884 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-474-068A-8

Query Match          54.6%; Score 2384; DB 2; Length 884;
Best Local Similarity 53.5%; Pred. No. 7e-213;
Matches 471; Conservative 123; Mismatches 221; Indels 66; Gaps 13;

QY 9 ASLLLLQV-CWLQCAASEP--CRAVFREAETLEAGAEQEPGQALGKV-FMGCPCQGEPA 64
DB 10 ALLLLQVSSWL-COELEPESCSGFSSEVYFPVPERHLERGHVLRGVFEGCTGRPT 68
QY 65 LFSTDNDFTVRNGETVQERRSLK----- 88
DB 69 AFFSEDSRFKATDGTITVKEHLKHLKLETSFLVRARDSSHRELSTKVTLKSMGHHHRH 128
QY 89 -----ERNP-LKIFPSKRI-LRRHKRDVWVAPISVPEKNGKGFPPORLNOLKSNKDRDT 139
DB 129 HHRDPAGESNELLMFFSVPLGRQRKRDWVIPPISCPENEKGEFPKNLVQIKSNDRKET 188
QY 140 KIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDREETIAKYELFQHAVSNGASVEDPM 199
DB 189 KVFYSITGQADKPEVGVGFIERETGLWKTQPLDREAIKAYILYSHAVSNGEVEDPM 248

```

```

QY 729 VLRNDVAPTIITPTMYRPRANPDEIGNFIENLKAANTDPTAPPYDTLLVFDYEGSSD 788
DB 785 T-RNDVAPTLMSVPOYRPRANPDEIGNFIENLKAADSDPTAPPYDLSLLVFDYEGSSG 843
QY 789 AASLSLSSASDQDDYDYLNEWGSRFKKLADMYGGGEDD 829
DB 844 AASLSLSSASDQDDYDYLNEWGSRFKKLADMYGGGEDD 884

RESULT 5
US-08-474-068A-8
; Sequence 8, Application US/08474068A
; Patent No. 5837525
; GENERAL INFORMATION:
; APPLICANT: Ranscht, Barbara
; TITLE OF INVENTION: T-Cadherin Adhesion Molecule
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESS: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/474, 068A
; APPLICATION NUMBER: US/08/474, 068A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/213,361
; FILING DATE: 14-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/607,293
; FILING DATE: 30-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1683
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 884 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-474-068A-8

```

```

Query Match          54.6%; Score 2384; DB 2; Length 884;
Best Local Similarity 53.5%; Pred. No. 7e-213;
Matches 471; Conservative 123; Mismatches 221; Indels 66; Gaps 13;

QY 9 ASLLLLQV-CWLQCAASEP--CRAVFREAETLEAGAEQEPGQALGKV-FMGCPCQGEPA 64
DB 10 ALLLLQVSSWL-COELEPESCSGFSSEVYFPVPERHLERGHVLRGVFEGCTGRPT 68
QY 65 LFSTDNDFTVRNGETVQERRSLK----- 88
DB 69 AFFSEDSRFKATDGTITVKEHLKHLKLETSFLVRARDSSHRELSTKVTLKSMGHHHRH 128
QY 89 -----ERNP-LKIFPSKRI-LRRHKRDVWVAPISVPEKNGKGFPPORLNOLKSNKDRDT 139
DB 129 HHRDPAGESNELLMFFSVPLGRQRKRDWVIPPISCPENEKGEFPKNLVQIKSNDRKET 188
QY 140 KIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDREETIAKYELFQHAVSNGASVEDPM 199
DB 189 KVFYSITGQADKPEVGVGFIERETGLWKTQPLDREAIKAYILYSHAVSNGEVEDPM 248

```



```
QY 200 NISIIIVTDNDHKPKFTQDFRGSVLGVLPGTSVMQVATDDEDDALITYNGVVAYSIHS 259
Db 249 EIVITVTDQNDNRPEFTQEFVGSVAEGVGSVMKVSATDADDVVNTVNAIAITVS 308
QY 260 QEPKOPDLMTIHRSTGTTSVSSGLDREKVPYELTIQATDMGSGSTTAVAVVEIL 319
Db 309 QDPPLPHKMTFVNRTDGTIVSLTSGLDRESYPTIIVVQAADLQEGSLTAKAVITVK 368
QY 320 DANDNAPMFDPOKYEAVHPNANVHEVQRLVTVDLAPNSPAMRATVYLINGDDGDGHFTI 379
Db 369 DINDNAPVFNSTYQGVPEVNEVARIATKLVTDADAENTPAWKVYTWVNDPDQFVV 427
QY 380 TTHPESNOGILTRKGLDFAKQHTLYVEVNEAPEVLKLPSTATVIVHVEDVNEAPV 439
Db 428 VTDPTNDGILKTAAGLDFAKQYILHVRVNEPEFSGSLVPSTATVIVDVNEAPI 487
QY 440 FVPSKVVEVQEGIPTGEPVCVYTAEDDPK-ENQKISYRILRDPAGWLAMPDPSGQVAV 498
Db 488 FMPAERREVEVPEDFGVQGEITSYTAREPDTFMDQKITVIRWDTANWLEINPETAIFTR 547
QY 499 GTLDREDEQFVRNNIYEVWVLAMONGSPPTGTGTLTLLTDVNDHGPVPEPQITICNQ 558
Db 548 AEMDREDAEHVKNSTYVALIATDGSPIATGTGTLTLLVLLVNDNAPPEPNWQFCQR 607
QY 559 SPVRHLNITDKLSPTSPTAQITDSDIYTAEVNEEG-DTVVLSKKFLKQDTYDV 617
Db 608 NPQPHIITLDPPLPNTSPFTAEITHGASVNWITIEYNDAAQESLILQPRKOLEIGEYKI 667
QY 618 HLSLSDHGNKQLTVIRATVCDCHGHVETCPGPKWGGF-----ILPVLGAVLALLF 724
Db 669 LLLVLLLVLRKRIKEPILLPEDDTRDNVYFVEEGGEGEDQDYDITQLHRGLEARPEV 728
QY 725 LILLLLLFLRRRTVVKPELLPPDDTRDNVYFVEEGGEGEDQDFDLSQLHRGLDARPEV 784
QY 729 VLNRDVAPTIPTPMYRPRANPDBEIGNFIENLKAANTDPTAPPYDITLLVDFYEGSGSD 788
Db 785 T-RNDVAPTLMSVQVPRPRANPDBEIGNFIENLKAANTDPTAPPYDITLLVDFYEGSGSD 843
QY 789 AASLSSLTSSASQDQDYDYLNEWGSFRKFLADMYGGEGEDD 829
Db 844 AASLSSLNSSSDQDQDYDYLNEWGNFRKFLADMYGGEGEDD 884

RESULT 6
US-08-472-481-7
; Sequence 7, Application US/08472481
; Patent No. 5863804
; GENERAL INFORMATION:
; APPLICANT: Ranscht, Barbara
; TITLE OF INVENTION: T-Cadherin Adhesion Molecule
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,481
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/213,361
; FILING DATE: 14-MAY-1994
; PRIOR APPLICATION DATA:
```

```
; APPLICATION NUMBER: US 07/607,293
; FILING DATE: 30-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1686
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 884 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-472-481-7

Query Match 54.6%; Score 2384; DB 2; Length 884;
Best Local Similarity 53.5%; Pred. No. 7a-213;
Matches 471; Conservative 123; Mismatches 221; Indels 66; Gaps 13;

QY 9 ASLLLLQV-CMLQCAASBP--CRAVFREAETLEAGGAEQFGQALGV-FWGCFGQBP 64
Db 10 ALLLLLLQVSSWL-CQELPESCSFGSSEVYTFVPVPERHLEGHVLRVREGCCTGRPT 68
QY 65 LFTSNDNFTVRNGSTVQERRSLK----- 88
Db 69 AFFSEDSRFKVATDGTITVGRHLKLHLETSFLVREARDSHRELSTKVTLKSMGHHHRH 128
QY 89 -----SRNP-LKIFPFSKRI-LRRHKEDWVAPISVPEKNGKPPQRLNOLKSNKDRDT 139
Db 129 HRDPASSNSPELLMFPSPYGLRQKRDWIPPLISCENEGKGFPPKMLVQIKSNRDET 188
QY 140 KIFYSITGFGADSPPEGVFAVEKETGWLLLNKPLDREIAKYELFEGHVAENGASVEDPM 199
Db 189 KVFYSITGOGADKPPGVGVFIIRERTGWLKVTOPLDREAIAKYILYSHAVSSNGEAVEDPM 248
QY 200 NISIIIVTDNDHKPKFTQDFRGSVLGVLPGTSVMQVATDDEDDALITYNGVVAYSIHS 259
Db 249 EIVITVTDQNDNRPEFTQEFVGSVAEGVGSVMKVSATDADDVVNTVNAIAITVS 308
QY 260 QEPKOPDLMTIHRSTGTTSVSSGLDREKVPYELTIQATDMGSGSTTAVAVVEIL 319
Db 309 QDPPLPHKMTFVNRTDGTIVSLTSGLDRESYPTIIVVQAADLQEGSLTAKAVITVK 368
QY 320 DANDNAPMFDPOKYEAVHPNANVHEVQRLVTVDLAPNSPAMRATVYLINGDDGDGHFTI 379
Db 369 DINDNAPVFNSTYQGVPEVNEVARIATKLVTDADAENTPAWKVYTWVNDPDQFVV 427
QY 380 TTHPESNOGILTRKGLDFAKQHTLYVEVNEAPEVLKLPSTATVIVHVEDVNEAPV 439
Db 428 VTDPTNDGILKTAAGLDFAKQYILHVRVNEPEFSGSLVPSTATVIVDVNEAPI 487
QY 440 FVPSKVVEVQEGIPTGEPVCVYTAEDDPK-ENQKISYRILRDPAGWLAMPDPSGQVAV 498
Db 488 FMPAERREVEVPEDFGVQGEITSYTAREPDTFMDQKITVIRWDTANWLEINPETAIFTR 547
QY 499 GTLDREDEQFVRNNIYEVWVLAMONGSPPTGTGTLTLLTDVNDHGPVPEPQITICNQ 558
Db 548 AEMDREDAEHVKNSTYVALIATDGSPIATGTGTLTLLVLLVNDNAPPEPNWQFCQR 607
QY 559 SPVRHLNITDKLSPTSPTAQITDSDIYTAEVNEEG-DTVVLSKKFLKQDTYDV 617
Db 608 NPQPHIITLDPPLPNTSPFTAEITHGASVNWITIEYNDAAQESLILQPRKOLEIGEYKI 667
QY 618 HLSLSDHGNKQLTVIRATVCDCHGHVETCPGPKWGGF-----ILPVLGAVLALLF 668
Db 669 LLLVLLLVLRKRIKEPILLPEDDTRDNVYFVEEGGEGEDQDYDITQLHRGLEARPEV 728
QY 725 LILLLLLFLRRRTVVKPELLPPDDTRDNVYFVEEGGEGEDQDFDLSQLHRGLDARPEV 784
QY 729 VLNRDVAPTIPTPMYRPRANPDBEIGNFIENLKAANTDPTAPPYDITLLVDFYEGSGSD 788
```

Db 785 T-RNDVAPLMSVQXRPANDEIGNFIDENLKAADSPTAPPYDLSLVDFYEGSGSE 843
 Qy 789 AASLSLTSASDQDQDYDYLNEWGSRFKKLADMYGGGDD 829
 Db 844 AASLSLTSASDQDQDYDYLNEWGSRFKKLADMYGGGDD 884

RESULT 7
 US-08-237-919-2
 ; Sequence 2, Application US/08237919
 ; Patent No. 5610281
 ; GENERAL INFORMATION:
 ; APPLICANT: Brenner, Michael B
 ; APPLICANT: Cepak, Karyn L
 ; TITLE OF INVENTION: Methods and Compositions for
 ; Modulating Heterotypic E-cadherin Interactions with T lymphocytes
 ; TITLE OF INVENTION: Modulating Heterotypic E-cadherin Interactions with T lymphocytes
 ; NUMBER OF SEQUENCES: 2
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: Wolf, Greenfield & Sacks P.C.
 ; STREET: 600 Atlantic Avenue
 ; CITY: Boston
 ; STATE: MA
 ; COUNTRY: USA
 ; ZIP: 02210
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/237,919
 ; FILING DATE:
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Plumer, Elizabeth R
 ; REGISTRATION NUMBER: 36,637
 ; REFERENCE/DOCKET NUMBER: B0801/7023
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 617-720-3500
 ; TELEFAX: 617-720-2441
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 878 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-237-919-2

Query Match 53.8%; Score 2348.5; DB 1; Length 878;
 Best Local Similarity 54.0%; Pred. No. 1.4e-209;
 Matches 472; Conservative 115; Mismatches 230; Indels 57; Gaps 12;

Qy 8 LASILLQVLCQAASEPCRAVREAEVTEAGAEQEPGQALGKV-FMGCPGQZ----- 62
 Db 8 LSGLLLLRSPLGQERSPPCLTRELHVHGAPEKRP--RLGRVNFEDCTGQRQTAI 65
 Qy 63 -----PALFSTDNDFTVRNGETVQERRSLK-----ERNP--- 92
 Db 66 FLTPPKVGTGDTVTRKPRFNPPTDPLGLRWDSYRFSKVTINTVGHHRPPPHQ 125
 Qy 93 -----LKIFF-SKILRRHRDVRVAPISYPENGKGPFPFQRLNQLKSNKDRDTKIF 142
 Db 126 ASVSGIQAEELLTFNSPGLRRQRKQDWIPPIPCPENEGKFPFKNLVQIKSNKDEKGVF 185
 Qy 143 YSITGPGADSPPEGVFAVEKTEHLLNKLKDREIAKYELFGHAYSENGASVEDPMNIS 202
 Db 186 YSITGCGADTPVGVFIETRTGKTEPLDRERIAITYLFGHAYSSNGAVEDPMNELL 245
 Qy 203 IIVTDQNDHKPFQDTFRGSLVSGVLPGTSVMQVATDDEDAIYTVNGVAYSIHSQEP 262
 Db 246 ITVTDQNDNKEPFTQEVFKGSMGALPGTSVMEVTAIDDDDDVNTYNAIAVTLISQDP 305

Qy 263 KDPHDLMTIHRSTGTISVISGLDREKVPYITITQATMDGDSGTTTAVAVVEILLDAN 322
 Db 306 ELFPDGMFTINRNTGIVSVVTTGLDRESFPYITLVVQAADLQSGSLSTTATAVITVTDN 365
 Qy 323 DNAPMFDQKYEAHVPENAVCHVEQRLVTDLDPNSPAMRATYILMGDDGDHFTITTH 382
 Db 366 DNPPIFNFTYKGVPEANEVNTTLTKVTDADAPNTPAMEAVYTIIL-NDGGGFVVTTN 424
 Qy 383 PESNOGILITRKGLDFFAKNQHTLYVEVNEAPFVCLKLPTSTATIVVHVEDVNEAPFVP 442
 Db 425 PVNNDGILKTAAGLDFFAKQYILHVAVTNVVPEVSLTTSTATVTVDVLDVNEGPFPV 484
 Qy 443 PSKVVEOEGIPTGEPVCVYTAEDPK-ENQKISYRILRDPAGWLAMPDPSQVYAVGTL 501
 Db 485 PEKRVSEDFGVGQBITSYTAQSPDTFMEQKITYRWRDTRNWLEINPDGTAISTRAEL 544
 Qy 502 DREDOFVRNNIYEVWLAMNGSPPTGTGTLLTLIDVNDHGVPEPEPQITICNOSP 561
 Db 545 DREDFEVRNNTYALIIATDNGSPVATGTGTLILSDVNDNAPIPEPRITFCERNPK 604
 Qy 562 RHLNITDKLSMTPSPQAOLTDSDIYTAEVNE-EGDTVLSLKKFLKQDYDVHLS 620
 Db 605 PQVINIHADLPNTSPFTELTGHRVFNWITQINDPTQESIIILKPMVALEVGDKYK 664
 Qy 621 LSDHGNKEQLTIRATVCDCHGVETC--PGPMKGGF---ILPVLGAVLALLFLLLVLL 674
 Db 665 LMDNQKQVTTLEVSVCDCGAGVCRKAQPVFAGLQIPAILGILGILALLLILLLL 724
 Qy 675 LLVRKKRIKEPLLPEDDTRDNVYFYEEGGEGEDQDYDTQLHRLGLEARPEVVLNDV 734
 Db 725 LFLRRRAVVKPELLPPEDDTRDNVYFYEEGGEGEDQDYDTQLHRLGLEARPEV 783
 Qy 735 APTIPTMYRPRANDEIGNFIENLKAADTPTAPPYDLSLVDFYEGSGSDAASLSS 794
 Db 784 APTLSPRYLPRANDEIGNFIENLKAADTPTAPPYDLSLVDFYEGSGSDAASLSS 843
 Qy 795 LTSSASDQDQDYDYLNEWGSRFKKLADMYGGGDD 828
 Db 844 LNSFSDXQDQDYDYLNEWGSRFKKLADMYGGGDD 877

RESULT 8
 US-08-732-429-2
 ; Sequence 2, Application US/08732429
 ; Patent No. 6300080
 ; GENERAL INFORMATION:
 ; APPLICANT:
 ; TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic
 ; TITLE OF INVENTION: E-cadherin Interactions with T Lymphocytes
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: Wolf, Greenfield & Sacks P.C.
 ; STREET: 600 Atlantic Avenue
 ; CITY: Boston
 ; STATE: MA
 ; COUNTRY: USA
 ; ZIP: 02210
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/732,429
 ; FILING DATE: herewith
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/237,919
 ; FILING DATE: 3 May 1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Plumer, Elizabeth R
 ; REGISTRATION NUMBER: 36,637
 ; REFERENCE/DOCKET NUMBER: B0801/7023
 ; TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 878 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-732-429-2

Query Match 53.8%; Score 2348.5; DB 4; Length 878;
Best Local Similarity 54.0%; Pred. No. 1.4e-209;
Matches 472; Conservative 115; Mismatches 230; Indels 57; Gaps 12;

```
QY 8 LASILLQVCLQCAASEPCRAVFEAEVTLQAGAEQEPQALGKV-FMGCPCQGE--- 62
DB 8 LSGILLLRSPGQSERSPPCLTRFLHVGAPAPPEKRP--RLGRVNFEDCTGRQTAI 65
QY 63 -----PALFSTNDNDFTVRNGETVQERRSLK-----ERNP--- 92
DB 66 FLTPIPKVGTDGVITVKRPLRFHNPDPFLGLRWDSYRKFSKVTLTNTVGHHRPPPHQ 125
QY 93 -----LKIFP-SKRILRRHKRDWVAPISVPENKGPFPORLNOLKSKNDRTKIF 142
DB 126 ASVSGIQAEILLTFPNSPGLRRQRDWIPIPCPENKGPFPKNLVQKSKNKEGKVF 185
QY 143 YSITPGADSPPEGVFAVEKETGMLLNKPLDREEIAKELFCHAVSNGASVEDPMNIS 202
DB 186 YSITGQADTPPVGVFIERTGMLKVTEPLDRERATYTLFHAVSSNGNAVEDPMEL 245
QY 203 IIVTDQNDHKPKTQDTPFGSVLEGVLPGTSVMQVATBDDDAIYTYNGVVAISHQEP 262
DB 246 IIVTDQNDKPEFTQEVFKGSVNEGALPGTSVMEVATDADDVNTYNAAIAYTILSDP 305
QY 263 KDPHLMFTIHRSTGTISVISGLDREKVPETLTQATDMGDGSGTTTAVAVEILDAN 322
DB 306 ELDPKMMFTINRNTGVISVVTGLDRESFPYTLVQAADLQEGSLSTATAVITVDTN 365
QY 323 DNAPMFDQKYEAHVPENAVCHVEQRLTVTDLPNPSPAWRATYLLIMGGDGDHFTITH 382
DB 366 DNPPIFNPTTKYQVPEANEANVITLKVTDADAPNTPAWEAVYTL-NDGGQCFVVTTN 424
QY 383 PESNOGILTRKGLDPEAKNOHTLYVEVNEAPFLKLTSTATIVVHVVDNEAPVFP 442
DB 425 PVNNDGILTKAGLDPEAKQYILHAVNTNVFVSLTSTATVTDVLDVNEGSIIFVP 484
QY 443 PSKVVEQEGIPTGEPCVVTAEADPK-ENOKISYRILRDPAGLAMPDPSGQVTAAGTL 501
DB 485 PEKRVSEDFGVQGEITSYTAQEPDTEMEQKITRYIRWDRNKLINPDTGAISRAEL 544
QY 502 DREDEQFVRNNIYEVWVLANDNGSPPTGTGTLILLDINDHGPVPEPRQITICNQSPV 561
DB 545 DREDFEHVKNSTYALIIATDNGSPVATGTGTLILLSDVNDNAPIPEPTIFFCERNPK 604
QY 562 RHLNITDKLSPTSPPFOAQLTDDSDIYWTAEVNE-EGDTVLSLKKFLKQDTYDVHLS 620
DB 605 PQVINIHADLPNTSPFFAELTHGRVPNKTIQYNDPTQESILLKPKMALEVDYKINLK 664
QY 621 LSHGNGEQLTVIRATVCDCHGHVETC--PGPWKGGF---ILPVGLAVALLFLLVLL 674
DB 665 LMDNQNKQDVTILEVSVCDCEGAAGVCRKAQPVVAGLQIPAILGILGLTALLILL 724
QY 675 LIVRKRKKEPILLPDDTRDNVYVYEGEGGEEDDYDTQLHGLEARPEVVLNDV 734
DB 725 LFLRRRAVKEPLLPEDDTRDNVYIYDEEGGEEDQDFDLQHLHGLDARPEVT-RNDV 783
QY 735 APTIIPMYRPRANPEIGNFIENLKAANTDPTAPPYDTLLVPDYEGSGSDAASLSS 794
DB 784 APTLMSVERVLPANPEIGNFIDENLKAADTPTAPPYDSILVFDYEGSGSEASLSS 843
QY 795 LTSASDQDQDYDYLNEWSRPFKLLADMYGGGED 828
DB 844 LNSSESQDQDYDYLNEWGNRFKLLADMYGGGED 877
```

RESULT 9

US-09-798-267-2
Sequence 2, Application US/09798267
Patent No. 6406870

GENERAL INFORMATION:

APPLICANT: Brenner, Michael
APPLICANT: Cepek, Karyn
TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic E-cadherin
FILE OF INVENTION: Interactions with T Lymphocytes
FILE REFERENCE: L0560/7008RP
CURRENT APPLICATION NUMBER: US/09/798,267
CURRENT FILING DATE: 2001-03-02
PRIOR APPLICATION NUMBER: US 08/237,919
PRIOR FILING DATE: 1994-05-03
PRIOR APPLICATION NUMBER: PCT/US 95/05518
PRIOR FILING DATE: 1995-05-03
PRIOR APPLICATION NUMBER: US 08/732,429
PRIOR FILING DATE: 1996-11-01
NUMBER OF SEQ ID NOS: 9
SOFTWARE: PatentIn version 3.0
SEQ ID NO 2
LENGTH: 878
TYPE: PRT
ORGANISM: Homo sapiens
US-09-798-267-2

Query Match 53.8%; Score 2348.5; DB 4; Length 878;
Best Local Similarity 54.0%; Pred. No. 1.4e-209;
Matches 472; Conservative 115; Mismatches 230; Indels 57; Gaps 12;

```
QY 8 LASILLQVCLQCAASEPCRAVFEAEVTLQAGAEQEPQALGKV-FMGCPCQGE--- 62
DB 8 LSGILLLRSPGQSERSPPCLTRFLHVGAPAPPEKRP--RLGRVNFEDCTGRQTAI 65
QY 63 -----PALFSTNDNDFTVRNGETVQERRSLK-----ERNP--- 92
DB 66 FLTPIPKVGTDGVITVKRPLRFHNPDPFLGLRWDSYRKFSKVTLTNTVGHHRPPPHQ 125
QY 93 -----LKIFP-SKRILRRHKRDWVAPISVPENKGPFPORLNOLKSKNDRTKIF 142
DB 126 ASVSGIQAEILLTFPNSPGLRRQRDWIPIPCPENKGPFPKNLVQKSKNKEGKVF 185
QY 143 YSITPGADSPPEGVFAVEKETGMLLNKPLDREEIAKELFCHAVSNGASVEDPMNIS 202
DB 186 YSITGQADTPPVGVFIERTGMLKVTEPLDRERATYTLFHAVSSNGNAVEDPMEL 245
QY 203 IIVTDQNDHKPKTQDTPFGSVLEGVLPGTSVMQVATBDDDAIYTYNGVVAISHQEP 262
DB 246 IIVTDQNDKPEFTQEVFKGSVNEGALPGTSVMEVATDADDVNTYNAAIAYTILSDP 305
QY 263 KDPHLMFTIHRSTGTISVISGLDREKVPETLTQATDMGDGSGTTTAVAVEILDAN 322
DB 306 ELDPKMMFTINRNTGVISVVTGLDRESFPYTLVQAADLQEGSLSTATAVITVDTN 365
QY 323 DNAPMFDQKYEAHVPENAVCHVEQRLTVTDLPNPSPAWRATYLLIMGGDGDHFTITH 382
DB 366 DNPPIFNPTTKYQVPEANEANVITLKVTDADAPNTPAWEAVYTL-NDGGQCFVVTTN 424
QY 383 PESNOGILTRKGLDPEAKNOHTLYVEVNEAPFLKLTSTATIVVHVVDNEAPVFP 442
DB 425 PVNNDGILTKAGLDPEAKQYILHAVNTNVFVSLTSTATVTDVLDVNEGSIIFVP 484
QY 443 PSKVVEQEGIPTGEPCVVTAEADPK-ENOKISYRILRDPAGLAMPDPSGQVTAAGTL 501
DB 485 PEKRVSEDFGVQGEITSYTAQEPDTEMEQKITRYIRWDRNKLINPDTGAISRAEL 544
QY 502 DREDEQFVRNNIYEVWVLANDNGSPPTGTGTLILLDINDHGPVPEPRQITICNQSPV 561
DB 545 DREDFEHVKNSTYALIIATDNGSPVATGTGTLILLSDVNDNAPIPEPTIFFCERNPK 604
QY 562 RHLNITDKLSPTSPPFOAQLTDDSDIYWTAEVNE-EGDTVLSLKKFLKQDTYDVHLS 620
```

```

Db 605 PQVINIHADLPNTSPTAEITLHGRVFNWNTIQNDPTQESIIKPKMALEVDYKINLK 664
Qy 621 LSHGKKEQITVIRATVCDCHGVETC--PGPWKGGF----ILPVLGAVLALLFLVL 674
Db 665 LMDNQKQVTTLEVSVCDEGAAGVCRKAQPVAGLQIPAILGILGILALLL 724
Qy 675 LLVRKKEKIEPILLPDDTRDNVYVYEGGGGEEODXDITOLHRLGLARBPWLNDV 734
Db 725 LFLRRRAVKEPLLPDDTRDNVYVYEGGGGEEODFDSQLHRLGLARPEVT-RNDV 783
Qy 735 APTIIPMYRPRANPDEIGNFIENLKAANTDPTAPPYDTLLVFDYEGSGDAASLS 794
Db 784 APTLMSVPRYLPRANPDEIGNFIDENLKAADTPTAPPYDLSLLVFDYEGSGEASLS 843
Qy 795 LTSSASQDQDYDYLNEWGSRFKKLADMYGGED 828
Db 844 LNSESDKDQDYDYLNEWGSRFKKLADMYGGED 877

```

```

RESULT 10
US-09-798-267-3
; Sequence 3, Application US/09798267
; Patent No. 6406870
; APPLICANT: Brenner, Michael
; APPLICANT: Cepek, Karen
; TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic E-cadherin
; FILE REFERENCE: L0560/7008ERP
; CURRENT APPLICATION NUMBER: US/09/798,267
; PRIOR FILING DATE: 2001-03-02
; PRIOR FILING DATE: 1994-05-03
; PRIOR FILING DATE: 1995-05-03
; PRIOR FILING DATE: 1998-11-01
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn.version 3.0
; SEQ ID NO 3
; LENGTH: 878
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-798-267-3

```

```

Query Match 53.8%; Score 2348.5; DB 4; Length 878;
Best Local Similarity 54.0%; Pred.No.1.4e-209;
Matches 472; Conservative 115; Mismatches 230; Indels 57; Gaps 12;

Qy 8 LASLILLOVLCQAAEPCEAVREAVTLEAGAEQPCQALGV-FMGCPQOE-----62
Db 8 LSGLLLLRLSPGLSQERSPPCLTLRELHVGAPAPPEKRP--RLGRVNFEDCTGRQTAI 65
Qy 63 -----PALFSTDDNDFVNRGETVQERRSLK-----ERNP---92
Db 66 FLTPPKVGTGDTVTKVKEPLRFNHTDPFLGLRWDSTYRKSTKVTLNTVGHRRPAPHQ 125
Qy 93 -----LKIPP-SKRLRRHKGDMVAPISVPENGKGPFPORLNQKSNKORDTKIF 142
Db 126 ASVSGIOAELLTFPNSPGLRQRDRQVIPPISCPENEGKGPFPKLVQIKSNKDKGKVF 185
Qy 143 YSITGPGADSPGEGVAVKETEGLLNKPLDREIAKYELFHAVENGASVEDPMNIS 202
Db 186 YSITGQGGADTPVGVFIERTGKLTWTEPLDRERIATYLFHAVSSNGNAVEDPMEIL 245
Qy 203 ITVDQNDHKPKFTQDTFRGVLGVLPGTSMVQVMTATDEDDALTYTNGVAVSIIHQEP 262
Db 246 ITVDQNDKPEFTQEVFKGSMEGALPGTSMVMTATDADDDVNTYNAAIATYLSQDP 305
Qy 263 KPDHDLMTIHRSGTISVSGGLDRKVPYTLTIQATMDGSGSTTTAVAVVEILDAN 322
Db 306 ELPDKNMTINRNGVSVVITGLDRESFPYTLTVQQAADLQEGELSTTATAVITVTDN 365

```

```

RESULT 11
PCI-US95-05518-2
; Sequence 2, Application PC/TUS9505518
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic
; E-cadherin Interactions with T Lymphocytes
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESS: Wolf, Greenfield & Sacks P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/05518
; FILING DATE: herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/237,919
; FILING DATE: 3 May 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R
; REGISTRATION NUMBER: 36,537
; REFERENCE/DOCKET NUMBER: B0801/7023
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:

```

```

Qy 323 DNAPMFDPKYEAHVPENAVGHEVQRLTVITDLDAPNSPAWRATYILMGDDGDGHFTTTH 382
Db 366 DNPFIENPTTKYQVPEANVWITLKVTDADAPNPAWEAVTIL-NDGQGGQFVVTTN 424
Qy 383 PESQGLITTRKGLDPAKQNTLYVETNEAPVFKLPTSTATVIVHVEDNEAPVFP 442
Db 425 PVNNDGILKTAKGLDFAKQYILHVAVTNVVPEVSLTSTATVTVVDVLDVNEGPFP 484
Qy 443 PSKVVEQEGPTGEPVCVTAEDPK-ENOKISYRILRDPAGWLAMDPSQGVTAAGTL 501
Db 485 PEKRVESVEDGVOQETSITAQEPDIFMEQKITYRWRDTRNWLINPDTGAISTRAEL 544
Qy 502 DREDSQFVRNNIYEVWLAMDNGSPPTGTGTLTLLTLDVNDHGFVPEPQITCNQSPV 561
Db 545 DREDPEHVKNTYATALLIATDNGSFVATGTGTLTLLIILSDVNDNAPIPEPRTIFCERNPK 604
Qy 562 RHLVNITDKLSBHTSPQAOLTDSDIYTAENB-EGDTVVLKFKKODTYDVHLS 620
Db 605 PQVINIHADLPNTSPTAEITLHGRVFNWNTIQNDPTQESIIKPKMALEVDYKINLK 664
Qy 621 LSHGKKEQITVIRATVCDCHGVETC--PGPWKGGF----ILPVLGAVLALLFLVL 674
Db 665 LMDNQKQVTTLEVSVCDEGAAGVCRKAQPVAGLQIPAILGILGILALLL 724
Qy 675 LLVRKKEKIEPILLPDDTRDNVYVYEGGGGEEODXDITOLHRLGLARBPWLNDV 734
Db 725 LFLRRRAVKEPLLPDDTRDNVYVYEGGGGEEODFDSQLHRLGLARPEVT-RNDV 783
Qy 735 APTIIPMYRPRANPDEIGNFIENLKAANTDPTAPPYDTLLVFDYEGSGDAASLS 794
Db 784 APTLMSVPRYLPRANPDEIGNFIDENLKAADTPTAPPYDLSLLVFDYEGSGEASLS 843
Qy 795 LTSSASQDQDYDYLNEWGSRFKKLADMYGGED 828
Db 844 LNSESDKDQDYDYLNEWGSRFKKLADMYGGED 877

```

Sequence 7, Application US/08474067
Patent No. 5811518
GENERAL INFORMATION:
APPLICANT: Rauscht, Barbara
TITLE OF INVENTION: T-cadherin Adhesion Molecule
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: United States
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,067
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER: US 08/213,361
FILING DATE: 14-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/607,293
FILING DATE: 30-OCT-1990
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 1682
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 837 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-474-067-7

Query Match 51.1%; Score 2348.5; DB 5; Length 878;
Best Local Similarity 54.0%; Pred. No. 1.4e-209;
Matches 472; Conservative 115; Mismatches 230; Indels 57; Gaps 12;
PCT-US95-05518-2

8 LASLLQLQVWLCQAASEPCRAVFEAEVTLFAGGAEPCQALGKV-FMGCPGQSE---62
8 LSGLLLLRPLSGQSESPPPCLTRHLVHGAPAPPEKRP-RLGRVNFEDCTGRQRTAI 65
63 -----PALFSTDDNDFTVRNGETVQERRSLK-----ERNP---92
66 FLTPIPKVGTDGVTIKRPLRFHNPDPFLGLRWSTYRKFSTKVTINTVGHHRPPPHQ 125
93 -----LKIRP-SKRILRRKRDWVAPISVPENKGFPPQRLNOLKSNKDRDTKIF 142
126 ASVSGIAQLTFPNSPGRLRQKRDWIPISCPENKGFPPKRLVQIKSNKDKGKVF 195
143 YSITPGADSPPEGVFAVEKETGMLLNKLPDREELAKYELFGHVSNGASVEDPMNIS 202
186 YSITGQADTPFVGVIETRETGLKVTPLDRERIATYLFHVSNGNAVEDDWEIL 245
203 IIVTDQNDHKPKFTQDTRFGSVLEGLVPGTSVMQVATDEDDAIYNGVVAISHSQEP 262
246 ITVTDQNDKPEFTQEVFKSGVMGALPGTSVMEVATDADDVNTNAIAIYTLISQDP 305
263 KDPHDLFTIHRSTGTISVISGLDREKVPPEYTLTQATDMGDGSTTTAVAVEILDAN 322
306 ELPDKMFTINRNTGVISVTTGLDRESFTYTLVQAAQLQEGSLSTTATAVITVDIN 365
323 DNAPMFDPKYEAHVPENAVGHVQRLVTVDLDAAPSNAWRYLLMGDDGDFHTITH 382
366 DNPPIFNFTYKGCVPNEANVITLKVTDADAPNTFAWEAVYTL-NDGCGQFVVTIN 424
383 PESNGGLITTRKGLDREKQHTLYEVETNEAPFVLKLPSTATIVVHVEDVNEAPVFP 442
425 FVNDGLKTKAGLDFEAKQYILHVAIVNVFVSEVLTSTATIVVDVLDNEGFIYP 484
443 PSKVEVEQEGIPTEPCVVTYABDPK-ENQKISYILRDPAGWLAMPDPSGQVTAAGTL 501
485 PEKRVESVEDFGVQSEITSYTAQEPDTFMEQKITRYIRWDRTNWLEINPDTGAISRAEL 544
502 DREDEQFVRNIIYEVWVLMNDGSPPTGTGTLILLTLIDVNDKGPVPEPQITICNSPV 561
545 DREDFEHVKNSTYALTIATDNGSPVATGTGTLILLLSVDVNDNAPPEPRTIFFCERNPK 604
562 RHVLNITDKLSPTSFPQALTDSDIYVTAEVNE-EGDTVVLSLKKFLKQDTYDVHLS 620
605 PQVINIHADLPNTSFTAEHLTHGRVPNNWTIQYNDPTQESILKPKMALEVGDYKINLK 664
621 LSDHNGEQLTVIRATVCDCHGVETC--PGPWKGF-----ILPVLGAVALLFLILLVL 674
665 LMDNQNDQVTTLEVSVCDCGAAGVCRKAQVPEAGLQIPAILGILGGLIALILLILL 724
675 LLVRKKEKIKPEILLPDDTRDNVYVYEGGEGEEDQYDITOLHGLEAPRPVIRNDV 734
725 LPIRRRAVKEPLLPEDDTRDNVYVYDEGGEGEEDQDFLSQLHGLDARPEVT-RNDV 783
735 APTIIPPMYRPRANPEITGNFIENLKAANTDPTAPPYDITLLVDFYEGSGDAASLS 794
784 APTLMSVPRVPRANPEIDGNFIDENLKAADTPTAPPYDLSLLVDFYEGSGSEASLS 843
795 LTSASDQDDYDYLNEWGSREKKLADMYGGED 828
844 LNSSESXQDDYDYLNEWGNRFKKLADMYGGED 877

RESULT 12
US-08-474-067-7

423 STATIVHVEDVNEAPVFPVSKVVEQVGPVGVVTAEDPKE-NOKISYRILRD 481
426 STASVLTVDVNEPPVFPVPIKRVGVPELVPVQVTSYTAEDPDRMRQKITRMSGD 485
482 PAGWLAMPDPSQVAVGTLDREDEQVRNIVYEVWVLANDNSPPTGTGTLILLIDV 541
486 PAGWYIHPENGIVTATQPDRESVHAI-NSTYKAILAVDNGIPDTTGTGTLILLIDV 544
542 NDHGFVPEPRQITTCNOSPVRHVLNITDKLSPHTSPFOAQLTDDSDIYTAEVNEEGDT 601
545 NDNGTPEPRSFECISQPEKQILSVKDLPHPTYPFKAALEHSGSSNNVTVEIRGQ-DE 603
502 VVLSLKFLKQDVTYDVHLSLDHGNKEQLTVIRATVCDCHGVETCP--GPWKGGF--- 655
604 LAMGLKKELEPGEYNI-FVKLTDSQKAVTQVKAQVCEGTAKNCERRSYIVGGLGVA 663
656 ILPVLGAVLALLFLLVLLLVKRRKI-KBPLLLPEDDTRDNVFFYEGEGGEDDQYDI 715
664 ILGILGILALLILLVLLLFARERKVEKEPLLPEDDMDRNVNVEEGGEDDQYDL 723
716 TOLHGLEARPEVLRNDVAPTIIPTMYRPRANPDEIGNFIENLKAANTDPTAPPYD 775
724 SQLHGLDARPEVI-RNDVAPPLMAAPQYRPRANPDEIGNFIENLKAANTDPTAPPYD 782
776 TLLVFDYEGSGSDAASLSLTSSTSSASDQDDYDYLNEWGSRFKKLADMYGGGEDD 829
783 SLLVFDYEGSGSEATSLSSLSNSASDQDDYDYLNEWGRFKKLAEYGGGEDD 836

RESULT 13

US-08-474-068A-7
; Sequence 7, Application US/08474068A
; Patent No. 5937525
; GENERAL INFORMATION:
; APPLICANT: Ranscht, Barbara
; TITLE OF INVENTION: T-Cadherin Adhesion Molecule
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,068A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/213,361
; FILING DATE: 14-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/607,293
; FILING DATE: 30-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1683
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 837 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-474-068A-7

Query Match 51.1%; Score 2231.5; DB 2; Length 837;
Best Local Similarity 54.0%; Pred. No. 1-le-198;
Matches 450; Conservative 117; Mismatches 212; Indels 55; Gaps 13;
QY 48 GOALGKV-FMGCPGQEPALF-STD-----NDQ-----FVTRNGETVQE 83
DB 6 GRELGRVSFACSGRPWAVYVPTDTRFKVNGGVVSNKRPILTYGRKISFTIYQAQDAMG 65
QY 84 RRS-----LKNRPULKI-PP--SKRILRRHQRDWWVADISVPEKNGK 122
DB 66 RRSARVTVGRHRRHRRHHHLLQDTPFAVLTFPPKHDFGRQRKRDWVIPPISLEHNRG 125
QY 123 PPOQLNOLKSKORDTKIFYSITGPAGDSPPEGVFAVEKETGMLLNKPLDRDREIAKYE 182
DB 126 PYPMLVQIKKNKESKVYISITGGADSPVGIFFIERETGHWLEVEQDLREKIDRYT 185
QY 183 LFGHAVSNGASVEDPMNISIIITDQNDHKPKFTQDTPRGSVLEGLVPGTSMVQVTADE 242
DB 186 LLSHAVSASGQPVEDPMBIIITVMDQNDKPVFIKEVFVGIENAKPGTSMVTVNATDA 245
QY 243 DDAIYTVNGVWAYS THSQEPKDPHDLMTIHRSTGTISVSSGLDREKVPETVTLTIQATD 302
DB 246 DAVNTDNGIVSIVSQPPRPHQPFTIDPAKGIISVLGTGLDRETTPNYTLIVQATD 305
QY 303 MDGSGTITTAVAVVEILDANDNAPMFPQKYEAHVPENAVGHVQRLTVTDLDAPNSPAW 362
DB 306 QEGKGLSNTATAIETVTDANDNIPFNPTMYEGVVEENKPGTEVARLTVTDQDAPGSPAW 365
QY 363 RATYLIMGGDGDHFTIITHPESNOGILTTKGLDFAKNQHTLYVVEVNEAPVFKLPT 422
DB 366 QAVYHIKSGNLGDAFSIITDESTNNGILKTAKGLDYETKSRDYLVVTVENKVPUSVITL 425
QY 423 STATIVHVEDVNEAPVFPVSKVVEQVGPVGVVTAEDPKE-NOKISYRILRD 481
DB 426 STASVLTVDVNEPPVFPVPIKRVGVPELVPVQVTSYTAEDPDRMRQKITRMSGD 485
QY 482 PAGWLAMPDPSQVAVGTLDREDEQVRNIVYEVWVLANDNSPPTGTGTLILLIDV 541
DB 486 PAGWYIHPENGIVTATQPDRESVHAI-NSTYKAILAVDNGIPDTTGTGTLILLIDV 544
QY 542 NDHGFVPEPRQITTCNOSPVRHVLNITDKLSPHTSPFOAQLTDDSDIYTAEVNEEGDT 601
DB 545 NDNGTPEPRSFECISQPEKQILSVKDLPHPTYPFKAALEHSGSSNNVTVEIRGQ-DE 603
QY 602 VVLSLKFLKQDVTYDVHLSLDHGNKEQLTVIRATVCDCHGVETCP--GPWKGGF--- 655
DB 604 LAMGLKKELEPGEYNI-FVKLTDSQKAVTQVKAQVCEGTAKNCERRSYIVGGLGVA 663
QY 656 ILPVLGAVLALLFLLVLLLVKRRKI-KBPLLLPEDDTRDNVFFYEGEGGEDDQYDI 715
DB 664 ILGILGILALLILLVLLLFARERKVEKEPLLPEDDMDRNVNVEEGGEDDQYDL 723
QY 716 TOLHGLEARPEVLRNDVAPTIIPTMYRPRANPDEIGNFIENLKAANTDPTAPPYD 775
DB 724 SQLHGLDARPEVI-RNDVAPPLMAAPQYRPRANPDEIGNFIENLKAANTDPTAPPYD 782
QY 776 TLLVFDYEGSGSDAASLSLTSSTSSASDQDDYDYLNEWGSRFKKLADMYGGGEDD 829
DB 783 SLLVFDYEGSGSEATSLSSLSNSASDQDDYDYLNEWGRFKKLAEYGGGEDD 836

RESULT 14

US-08-472-481-6
; Sequence 6, Application US/08472481
; Patent No. 5863804
; GENERAL INFORMATION:
; APPLICANT: Ranscht, Barbara
; TITLE OF INVENTION: T-Cadherin Adhesion Molecule
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700

542 NDHGFVPEPRQITICNQSPVRHVLNITKDLSPHTSPFQAOLTDSDIYTAENVBEGDT 601
545 NDNGPTPEPRSGFEICSRQPEKQILSVKDLPHPTYPFKAALEHGSSNNWTVIRGQ-DE 603
602 VVLSLKKEFKODTYDVHLSDHGNKEQLTWIRATVCDCHGHVETCP--GPWKGF----- 655
604 LANGLKSELEGEYNIFVKLTDSQGAQVTKAQVCEGTAKNCERSSRSTVIGLGVPA 663
656 ILPVLGAVLALIFLLVLLVLRKKRKEPLLLPEDDTRDNVFFYEGEGGEEEDQDYDI 715
664 ILGILGGILALLLILLLLFARRKKEPEPLLPEDDVRNVVNYDEEGGEEEDQDYDL 723
716 TOLHRLGARBEVVLNDVATIIPTMYRRPRANPDEIGNFIENLKAANTDPTAPYD 775
724 SQLHGLDARPEVI-RNDVAPPLMAAPQYRPRANPDEIGNFIENLKAANTDPTAPYD 782
776 TLLVDEYEGSGSAAASLSLTSASDQDQDYDYNLWGSFRKFLADMYGGGDD 829
783 SLLVDFYEGGSEATSLSSLNSSASDQDQDYDYNLWGNRFRKFLAELXGGGDD 836

RESULT 15
US-09-417-039-11
; Sequence 11, Application US/09417039A
; Patent No. 6485972
; GENERAL INFORMATION:
; APPLICANT: McMahon, Andrew P.
; APPLICANT: Parr, Brian A.
; APPLICANT: Vaino, Seppo
; TITLE OF INVENTION: WNT SIGNALING IN REPRODUCTIVE ORGANS
; FILE REFERENCE: 00246/232001
; CURRENT APPLICATION NUMBER: US/09/417,039A
; CURRENT FILING DATE: 1999-10-12
; EARLIER APPLICATION NUMBER: US 60/109,355
; EARLIER FILING DATE: 1998-10-15
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 906
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-417-039-11

Query Match 39.2%; Score 1714; DB 4; Length 906;
Best Local Similarity 42.8%; Pred. No. 2,3e-150;
Matches 391; Conservative 129; Mismatches 291; Indels 102; Gaps 27;

542 NDHGFVPEPRQITICNQSPVRHVLNITKDLSPHTSPFQAOLTDSDIYTAENVBEGDT 601
545 NDNGPTPEPRSGFEICSRQPEKQILSVKDLPHPTYPFKAALEHGSSNNWTVIRGQ-DE 603
602 VVLSLKKEFKODTYDVHLSDHGNKEQLTWIRATVCDCHGHVETCP--GPWKGF----- 655
604 LANGLKSELEGEYNIFVKLTDSQGAQVTKAQVCEGTAKNCERSSRSTVIGLGVPA 663
656 ILPVLGAVLALIFLLVLLVLRKKRKEPLLLPEDDTRDNVFFYEGEGGEEEDQDYDI 715
664 ILGILGGILALLLILLLLFARRKKEPEPLLPEDDVRNVVNYDEEGGEEEDQDYDL 723
716 TOLHRLGARBEVVLNDVATIIPTMYRRPRANPDEIGNFIENLKAANTDPTAPYD 775
724 SQLHGLDARPEVI-RNDVAPPLMAAPQYRPRANPDEIGNFIENLKAANTDPTAPYD 782
776 TLLVDEYEGSGSAAASLSLTSASDQDQDYDYNLWGSFRKFLADMYGGGDD 829
783 SLLVDFYEGGSEATSLSSLNSSASDQDQDYDYNLWGNRFRKFLAELXGGGDD 836

CITY: San Diego
STATE: California
COUNTRY: United States
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/472,481
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/213,361
FILING DATE: 14-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/607,293
FILING DATE: 30-OCT-1990
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 1686
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8943
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 837 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-472-481-6

Query Match 51.1%; Score 2231.5; DB 2; Length 837;
Best Local Similarity 54.0%; Pred. No. 1.1e-198;
Matches 450; Conservative 117; Mismatches 212; Indels 55; Gaps 13;

48 GOALGKV-FMGCPQOEALP-STD-----NDD-----PTVRNGETVQS 83
6 GRELGVSFAACGRPWAVTPTDFKNGDGVSTYKRLTYLGRKISFTIYAQAMGK 65
84 RRS-----LKERNFLKI-FP--SKRIILRRHRKRWVAVIPSPENGK 122
66 RHSGARVTVGRHRRHRRHHHLQDTPAVLTFPKHDPGLRRQKRWVIPPISLENHRG 125
123 PFPORNLQNSKDRDTKPIYSITGQADSPPGGFAVEKETGWLKLLNKLDRREBIAKYE 182
126 PYPMLVQIKSNKDKESKVTYSITGQADSPPGGFAVEKETGWLKLLNKLDRREBIAKYE 185
183 LFGHAVSENGASVEDPNIIGIITVDQNDHKPKFTQDTFRGSVLEGVLPGTSVMQVTAIDE 242
186 LLSHAVSASQVDEPMEIITVMDQNDKKEVFIKEVFGYIEENAKPGTSVMVNAIDA 245
243 DDALYTVNGVAVSIHSEQPKDHLMTFTIHRSTGTISVSSGLDRKVKPEYTLTIQATD 302
246 DDAVNTDNGIVSVISVQCPRPFPQFTIDPAKGIISVLGTGLDRETTENYTLIQAATD 305
303 MDGSGSTTAVAVVEILDANDNAPMFPQKYEAVHPENAVGHEVQRLTVDLDAPNSPAW 362
306 QEGKGLSNTAIIIEVDANDNPIFNPTNYGVEENKEGTEVARKLTVDQAPGSPAW 365
363 RAYTLINGGDDGHFTITTHPESNGGLITRKGDLDFEAKNQHTLYVEVTNEAPFLKLPT 422
366 QAVYHIKSGNLGDAFSIITDPTSTNNGILKTAAGLDYETKSRDYLVTVVENKVPISPIIL 425
423 STAIIVHVEVDNEAPVFPSPKVEVQEGIPTEGVCVTAEDPQKE-NOKISYRLIRD 481
426 STASVLTVLDVNEPFPVFPKVGVPEDLPVGCQVTSYTABDFDRMKFKITYRMGSD 485
482 PAGMLANDPDSQGVATVGTGLDREDEQFVRNNIYEMVWMLAMDNPSPTTGTGTLILLIDV 541
486 PAGWLVIHPENGIVATQPLDRESVHAI-NSTYKAILAVDNGIPDTGTGTLILLIDV 544

Db 364 NTATAVITVTDVNDNPEFTAMTYGEVPENRVDIIIVANLTVTDKQOPHTPAMNAVYRIS 423
 Qy 370 GGDGDHFTIITHPESNOGILTRKGLDFAKNQOHTLYVEVTNEAPEVLKL---PTSTAT 426
 Db 424 GGDPTGRFAIQTDENSNDGLVTVVKPIDFETNRMFVLTVAAENQVPLAKGIGHPPQSTAT 483
 Qy 427 IVWVEDVNEAPVFPVPSKVVEOEGIPTGEPVCVYTAEDDDK-ENQKISYRILRDPAGW 485
 Db 484 VSVTVIDVNEPYPAPNPKIIRQEEGLHAGTMLTFTTAQDPDRYMQQNIHRTKLSPPANW 543
 Qy 486 LAMPDSCQVAVGTLDREDEBQFVRNNIYEVWVLAMDNQSPPTGTGCTLLTLTLDVNDHG 545
 Db 544 LKIDPVNGQITTIIVLDRESPN-VKNNIYNATFLASDNQIPPMMSGTGTLOIYLLDINDNA 602
 Qy 546 PVPEPRQITICNQSPVRHVLNIT--DKDLSPHTSPPFOQLTDDSDIY---WTAEVNEEGD 600
 Db 603 PQVLFQEAETC-ETFPDPSNITALDYDIDFNAGFFAFDLPLSPVTIKRNWTI-TRLNGD 660
 Qy 601 TVVLSLK-KPLKQDTYDVHLSLSDHGN--KEQLTVIRATVCDCHG-----VETCPGPWK 652
 Db 661 FAQLNLKIKFLEAGIYEVPIIITDSGNPPKSNISILRVKVCQDSNGDCTDVRIVGAGL 720
 Qy 653 GGFILPVGLAVLALLFLLVLLLV-----RKGRKIKPELLLPEDDTRDNVFFYG 702
 Db 721 G-----TGAIITALLCIIILLILVLMFVVMKRDKERQAKQLLIDPEDDVRDNILKYD 774
 Qy 703 BEGGGEEDDYDITQLHGLEAREPV-----VLENDVAPTIIPTMYRPERA--NPDEIG 755
 Db 775 BEGGGEEDQDYLSQLQOPDTVEPDAIKPVGIRMDERP-IHAEPQYFVRSAAHPHGDIG 833
 Qy 756 NFIIENLKAANTDPTAPPYDTLLVFDYEGSGSDAASLSLTSSASDQDQDYDYLNEWGSR 815
 Db 834 DFINEGLKAADNDPTAPPYDLSLVFDYEGSGSTAGSLSSLSNSSSGGGEQDYDYLNDWGP 893
 Qy 816 FKLLADMYGGGED 828
 Db 894 FKLLADMYGGGDD 906

Search completed: September 21, 2004, 22:10:10
 Job time : 36 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 22:09:14 ; Search time 143 Seconds
(without alignments)
1861.675 Million cell updates/sec

Title: US-09-916-849A-1
Perfect score: 4369
Sequence: 1 MGLPRGLASLLQLQVCWLQ.....NEWGSRFKLADMYGGEDD 829

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1342398 seqs, 321133274 residues

Total number of hits satisfying chosen parameters: 1342398

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/prodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/prodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/prodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/prodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/prodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/prodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/prodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/prodata/2/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/prodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4369	100.0	829	9	US-09-905-983-50
2	4369	100.0	829	10	US-09-916-849A-1
3	4369	100.0	829	10	US-09-975-723A-1
4	4369	100.0	829	12	US-10-058-270A-126
5	4369	100.0	829	14	US-10-174-677-21
6	4369	100.0	829	14	US-10-158-123-1
7	4369	100.0	829	15	US-10-295-027-783
8	4369	100.0	829	15	US-10-295-027-800
9	4369	100.0	829	15	US-10-295-027-863
10	4369	100.0	829	15	US-10-295-027-896
11	4361	99.8	829	12	US-10-329-345-16
12	4361	99.8	829	12	US-10-274-177-16
13	4361	99.8	829	14	US-10-301-822-18
14	4361	99.8	829	15	US-10-295-027-338
15	4361	99.8	829	15	US-10-295-027-1246

16	4361	99.8	829	15	US-10-087-080-14	Sequence 14, Appl
17	4361	99.8	829	16	US-10-650-112-16	Sequence 16, Appl
18	2420	55.4	882	14	US-10-097-340-37	Sequence 37, Appl
19	2388	54.7	899	9	US-09-905-983-5	Sequence 5, Appl
20	2348.5	53.8	878	13	US-10-165-049-2	Sequence 2, Appl
21	2348.5	53.8	878	13	US-10-165-049-3	Sequence 3, Appl
22	2341.5	53.6	878	9	US-09-905-983-48	Sequence 48, Appl
23	2341.5	53.6	878	15	US-10-173-551-34	Sequence 34, Appl
24	2285	52.3	894	12	US-10-425-114-39106	Sequence 39106, A
25	2216.5	50.7	821	14	US-10-174-677-28	Sequence 28, Appl
26	1714	39.2	906	14	US-10-174-677-19	Sequence 19, Appl
27	1707	39.1	906	9	US-09-905-983-46	Sequence 46, Appl
28	1707	39.1	906	14	US-10-177-293-43	Sequence 43, Appl
29	1707	39.1	906	16	US-10-778-146-6	Sequence 6, Appl
30	1704.5	39.0	912	9	US-09-905-983-2	Sequence 2, Appl
31	1702	39.0	906	9	US-09-746-491-46	Sequence 46, Appl
32	1692.5	38.7	912	9	US-09-746-491-49	Sequence 49, Appl
33	1685.5	38.6	906	9	US-09-746-491-48	Sequence 48, Appl
34	1679	38.4	877	9	US-09-746-491-47	Sequence 47, Appl
35	1648.5	37.7	916	15	US-10-295-027-286	Sequence 286, Appl
36	1638	37.5	824	15	US-10-104-047-2515	Sequence 2515, Ap
37	1621.5	37.1	916	14	US-10-174-677-22	Sequence 22, Appl
38	1621.5	37.1	916	15	US-10-116-275-201	Sequence 201, App
39	1422	32.5	848	12	US-09-805-020-56	Sequence 56, Appl
40	1321.5	30.2	814	14	US-10-174-677-14	Sequence 14, Appl
41	1242	28.4	819	15	US-10-161-493-62	Sequence 62, Appl
42	1164	26.6	556	10	US-09-880-573-98	Sequence 98, Appl
43	1030.5	23.6	785	14	US-10-174-677-25	Sequence 25, Appl
44	1021.5	23.4	713	14	US-10-174-677-13	Sequence 13, Appl
45	1021.5	23.4	713	15	US-10-408-765A-394	Sequence 394, App

ALIGNMENTS

RESULT 1

US-09-905-983-50
; Sequence 50, Application US/0905983
; Patent No. US20020045591A1
; GENERAL INFORMATION:
; APPLICANT: Geiger, Benjamin
; APPLICANT: Ben-Ze'ev, Avri
; APPLICANT: Sadot, Einat
; TITLE OF INVENTION: METHODS AND THERAPEUTIC COMPOSITIONS FOR TREATING CANCER
; FILE REFERENCE: 01/22326
; CURRENT APPLICATION NUMBER: US/09/905,983
; CURRENT FILING DATE: 2001-09-28
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-905-983-50

Query Match	100.0%	Score 4369;	DB 9;	Length 829;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 829;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MGLPRGLASLLQLQVCWLQCAASEPCRAVPREAVTLEAGGAOEPCGALGKVFMGCPG	60	
Db	1	MGLPRGLASLLQLQVCWLQCAASEPCRAVPREAVTLEAGGAOEPCGALGKVFMGCPG	60	
QY	61	QEPALFSTNDTDFVRNGETVOERSLKERNPLKIPPSKRIILRRHKRDVWVAPISVPENG	120	
Db	61	QEPALFSTNDTDFVRNGETVOERSLKERNPLKIPPSKRIILRRHKRDVWVAPISVPENG	120	
QY	121	KGPFPPQRLNQLSKNDRDTKIFYSITGPGADSPPEGVFAVEKETGWLINLKPDLDEEIAK	180	
Db	121	KGPFPPQRLNQLSKNDRDTKIFYSITGPGADSPPEGVFAVEKETGWLINLKPDLDEEIAK	180	
QY	181	YELPHGVHENGASVEDPMNIIIVTDONDHKPKTQDTFRGSLVLEGLVPSVNCVTAT	240	

```

Db 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSVLEGLVPGTSVMQVAT 240
QY 241 DEDDAIYTYNGVAVYIHSQBPQKPHDLMTIHRSTGTISVISSGLDREKVPYTLTQIA 300
Db 241 DEDDAIYTYNGVAVYIHSQBPQKPHDLMTIHRSTGTISVISSGLDREKVPYTLTQIA 300
QY 301 TMDGCGSTTTAVAVVEILDANDNAPMDPKQYEAHVPENAVHVEVQRLTVDLDAVNSP 360
Db 301 TMDGCGSTTTAVAVVEILDANDNAPMDPKQYEAHVPENAVHVEVQRLTVDLDAVNSP 360
QY 361 AWRATYLMGDDGDDHFTIITHPESNOGILTRKGLDPEAKNQHTLVVETNEAPFVLK 420
Db 361 AWRATYLMGDDGDDHFTIITHPESNOGILTRKGLDPEAKNQHTLVVETNEAPFVLK 420
QY 421 PTSTATIIVHVEDVNEAPVFPVPPSKVVEVQSGIPTGEPVCVYTAEDPDKENQKISYILR 480
Db 421 PTSTATIIVHVEDVNEAPVFPVPPSKVVEVQSGIPTGEPVCVYTAEDPDKENQKISYILR 480
QY 481 DPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVVWLAMDNGSPPTTGTGLLLTLD 540
Db 481 DPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVVWLAMDNGSPPTTGTGLLLTLD 540
QY 541 VNDHGPVPEPRQITICNOSPVRHVLNITDKLSPTSPFQAQLTDDSDIYWTAEVNEEGD 600
Db 541 VNDHGPVPEPRQITICNOSPVRHVLNITDKLSPTSPFQAQLTDDSDIYWTAEVNEEGD 600
QY 601 TVVLSLKKFLKQDQYDVHLSLSDHGNKEQLTVIRATVCDCHGVETCPGPKWGGFILPVL 660
Db 601 TVVLSLKKFLKQDQYDVHLSLSDHGNKEQLTVIRATVCDCHGVETCPGPKWGGFILPVL 660
QY 661 CAVLALLFLLVLLVLLVLRKKKIKKEPILLPDDTRDNVFFYVYGGEGEEDQDIDITQLHR 720
Db 661 CAVLALLFLLVLLVLLVLRKKKIKKEPILLPDDTRDNVFFYVYGGEGEEDQDIDITQLHR 720
QY 721 GLEARPEVVLNDVAPTIPTPMYRPRPANDPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
Db 721 GLEARPEVVLNDVAPTIPTPMYRPRPANDPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
QY 781 DYEGSGDAASLSLTSASDQDQDYDLNWSRFRFKKLADMYGGEDD 829
Db 781 DYEGSGDAASLSLTSASDQDQDYDLNWSRFRFKKLADMYGGEDD 829

RESULT 2
US-09-916-849A-1
; Sequence 1, Application US/09916849A
; Publication No. US20030086934A1
; GENERAL INFORMATION:
; APPLICANT: Bostein, et al.
; TITLE OF INVENTION: Basal Markers in Breast Cancer and Related Reagents
; FILE OF INVENTION: Uses Thereof
; FILE REFERENCE: 2002850-0024
; CURRENT APPLICATION NUMBER: US/09/916,849A
; CURRENT FILING DATE: 2001-07-26
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 829
; TYPE: PR1
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Sequence of
; OTHER INFORMATION: Cadherin 3
US-09-916-849A-1

Query Match 100.0%; Score 4369; DB 10; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPRGLASLLLLQVCWLQCAASEPCRAVFAEVTLEAGGAEQBPFGALGVFMGCPG 60
Db 1 MGLPRGLASLLLLQVCWLQCAASEPCRAVFAEVTLEAGGAEQBPFGALGVFMGCPG 60

```

```

QY 61 QBPALFSTNDNDDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDWVAPISVPENG 120
Db 61 QBPALFSTNDNDDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDWVAPISVPENG 120
QY 121 KGPFFQRLNQLSKNDKRDTKIFYSITGPGADSPPGVFAVEKETGWLKLLNKLPLDREETA 180
Db 121 KGPFFQRLNQLSKNDKRDTKIFYSITGPGADSPPGVFAVEKETGWLKLLNKLPLDREETA 180
QY 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSVLEGLVPGTSVMQVAT 240
Db 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSVLEGLVPGTSVMQVAT 240
QY 241 DEDDAIYTYNGVAVYIHSQBPQKPHDLMTIHRSTGTISVISSGLDREKVPYTLTQIA 300
Db 241 DEDDAIYTYNGVAVYIHSQBPQKPHDLMTIHRSTGTISVISSGLDREKVPYTLTQIA 300
QY 301 TMDGCGSTTTAVAVVEILDANDNAPMDPKQYEAHVPENAVHVEVQRLTVDLDAVNSP 360
Db 301 TMDGCGSTTTAVAVVEILDANDNAPMDPKQYEAHVPENAVHVEVQRLTVDLDAVNSP 360
QY 361 AWRATYLMGDDGDDHFTIITHPESNOGILTRKGLDPEAKNQHTLVVETNEAPFVLK 420
Db 361 AWRATYLMGDDGDDHFTIITHPESNOGILTRKGLDPEAKNQHTLVVETNEAPFVLK 420
QY 421 PTSTATIIVHVEDVNEAPVFPVPPSKVVEVQSGIPTGEPVCVYTAEDPDKENQKISYILR 480
Db 421 PTSTATIIVHVEDVNEAPVFPVPPSKVVEVQSGIPTGEPVCVYTAEDPDKENQKISYILR 480
QY 481 DPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVVWLAMDNGSPPTTGTGLLLTLD 540
Db 481 DPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVVWLAMDNGSPPTTGTGLLLTLD 540
QY 541 VNDHGPVPEPRQITICNOSPVRHVLNITDKLSPTSPFQAQLTDDSDIYWTAEVNEEGD 600
Db 541 VNDHGPVPEPRQITICNOSPVRHVLNITDKLSPTSPFQAQLTDDSDIYWTAEVNEEGD 600
QY 601 TVVLSLKKFLKQDQYDVHLSLSDHGNKEQLTVIRATVCDCHGVETCPGPKWGGFILPVL 660
Db 601 TVVLSLKKFLKQDQYDVHLSLSDHGNKEQLTVIRATVCDCHGVETCPGPKWGGFILPVL 660
QY 661 CAVLALLFLLVLLVLLVLRKKKIKKEPILLPDDTRDNVFFYVYGGEGEEDQDIDITQLHR 720
Db 661 CAVLALLFLLVLLVLLVLRKKKIKKEPILLPDDTRDNVFFYVYGGEGEEDQDIDITQLHR 720
QY 721 GLEARPEVVLNDVAPTIPTPMYRPRPANDPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
Db 721 GLEARPEVVLNDVAPTIPTPMYRPRPANDPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
QY 781 DYEGSGDAASLSLTSASDQDQDYDLNWSRFRFKKLADMYGGEDD 829
Db 781 DYEGSGDAASLSLTSASDQDQDYDLNWSRFRFKKLADMYGGEDD 829

RESULT 3
US-09-975-723A-1
; Sequence 1, Application US/09975723A
; Publication No. US20030108529A1
; GENERAL INFORMATION:
; APPLICANT: Focy, Ramsey
; APPLICANT: Nackman, Gary
; TITLE OF INVENTION: Improvement of Endothelial Cell-Cell
; FILE OF INVENTION: Cohesion
; FILE REFERENCE: 601-1-101N
; CURRENT APPLICATION NUMBER: US/09/975,723A
; CURRENT FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: 60/241,216
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/243,693
; PRIOR FILING DATE: 2000-10-27
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 829

```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-975-723A-1

Query Match      100.0%; Score 4369; DB 10; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGLPRGLASLLQLQVLCWQCAASPCRAVFEAEVLEAGGAEOBPGQALGVFGCPG 60
Db 1 MGLPRGLASLLQLQVLCWQCAASPCRAVFEAEVLEAGGAEOBPGQALGVFGCPG 60

Qy 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIPPSKRIILRRHKRDWVAPISVPENG 120
Db 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIPPSKRIILRRHKRDWVAPISVPENG 120

Qy 121 KGPPQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWLINLKPDRREIAK 180
Db 121 KGPPQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWLINLKPDRREIAK 180

Qy 121 KGPPQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWLINLKPDRREIAK 180
Db 121 KGPPQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWLINLKPDRREIAK 180

Qy 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSLVLEGLPGTSVMQVAT 240
Db 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSLVLEGLPGTSVMQVAT 240

Qy 241 DEDDAIYTYNGVAVSIHSGEPKDPHDLMTTIHRSTGTISVISGLDREKVPYTLTIOA 300
Db 241 DEDDAIYTYNGVAVSIHSGEPKDPHDLMTTIHRSTGTISVISGLDREKVPYTLTIOA 300

Qy 241 DEDDAIYTYNGVAVSIHSGEPKDPHDLMTTIHRSTGTISVISGLDREKVPYTLTIOA 300
Db 241 DEDDAIYTYNGVAVSIHSGEPKDPHDLMTTIHRSTGTISVISGLDREKVPYTLTIOA 300

Qy 301 TMDGDSSTTTAVAVVEILDANDNAPMDFPKYEAHVPENAVGEVORLTVTDLDAVNSP 360
Db 301 TMDGDSSTTTAVAVVEILDANDNAPMDFPKYEAHVPENAVGEVORLTVTDLDAVNSP 360

Qy 301 TMDGDSSTTTAVAVVEILDANDNAPMDFPKYEAHVPENAVGEVORLTVTDLDAVNSP 360
Db 301 TMDGDSSTTTAVAVVEILDANDNAPMDFPKYEAHVPENAVGEVORLTVTDLDAVNSP 360

Qy 361 AWRATYIMGGDDGDHFTTIHPESNQILITTRKGLDPEAKNQHTLYVEVNEAPFVLK 420
Db 361 AWRATYIMGGDDGDHFTTIHPESNQILITTRKGLDPEAKNQHTLYVEVNEAPFVLK 420

Qy 421 PTSTATIIVHVEDVNEAPVFPVPSKVVEQEGIPTEGVCVYTAEDPKENOKISYILR 480
Db 421 PTSTATIIVHVEDVNEAPVFPVPSKVVEQEGIPTEGVCVYTAEDPKENOKISYILR 480

Qy 421 PTSTATIIVHVEDVNEAPVFPVPSKVVEQEGIPTEGVCVYTAEDPKENOKISYILR 480
Db 421 PTSTATIIVHVEDVNEAPVFPVPSKVVEQEGIPTEGVCVYTAEDPKENOKISYILR 480

Qy 481 DPAGWLAMPDPSGQVTAAGTLDREDEQFVRNNIYEVWVLAMDNQSPPTTGTGLTLTLLID 540
Db 481 DPAGWLAMPDPSGQVTAAGTLDREDEQFVRNNIYEVWVLAMDNQSPPTTGTGLTLTLLID 540

Qy 541 VNDHGPVPEPRQITICNQSPVRHVLNITDKDLSPTSPFQALTDSDIYWTAEVNEEGD 600
Db 541 VNDHGPVPEPRQITICNQSPVRHVLNITDKDLSPTSPFQALTDSDIYWTAEVNEEGD 600

Qy 601 TVVLSLKKFLKQDVTYDVHLSLSDHGKKEQLTVIRATVCDCHGHVETCPGPKGGFLLPVL 660
Db 601 TVVLSLKKFLKQDVTYDVHLSLSDHGKKEQLTVIRATVCDCHGHVETCPGPKGGFLLPVL 660

Qy 601 TVVLSLKKFLKQDVTYDVHLSLSDHGKKEQLTVIRATVCDCHGHVETCPGPKGGFLLPVL 660
Db 601 TVVLSLKKFLKQDVTYDVHLSLSDHGKKEQLTVIRATVCDCHGHVETCPGPKGGFLLPVL 660

; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Breast Cancer, Compositions and
; FILE REFERENCE: 018501-005210US
; CURRENT APPLICATION NUMBER: US/10/058,270A
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: US 60/263,965
; PRIOR FILING DATE: 2001-01-24
; PRIOR APPLICATION NUMBER: US 60/265,928
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: US 09/829,472
; PRIOR FILING DATE: 2001-04-09
; PRIOR APPLICATION NUMBER: US 60/282,698
; PRIOR FILING DATE: 2001-04-09
; PRIOR APPLICATION NUMBER: US 60/288,590
; PRIOR FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: US 60/294,443
; PRIOR FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 141
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 126
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-058-270A-126

Query Match      100.0%; Score 4369; DB 12; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGLPRGLASLLQLQVLCWQCAASPCRAVFEAEVLEAGGAEOBPGQALGVFGCPG 60
Db 1 MGLPRGLASLLQLQVLCWQCAASPCRAVFEAEVLEAGGAEOBPGQALGVFGCPG 60

Qy 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIPPSKRIILRRHKRDWVAPISVPENG 120
Db 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIPPSKRIILRRHKRDWVAPISVPENG 120

Qy 121 KGPPQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWLINLKPDRREIAK 180
Db 121 KGPPQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWLINLKPDRREIAK 180

Qy 121 KGPPQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWLINLKPDRREIAK 180
Db 121 KGPPQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWLINLKPDRREIAK 180

Qy 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSLVLEGLPGTSVMQVAT 240
Db 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSLVLEGLPGTSVMQVAT 240

Qy 241 DEDDAIYTYNGVAVSIHSGEPKDPHDLMTTIHRSTGTISVISGLDREKVPYTLTIOA 300
Db 241 DEDDAIYTYNGVAVSIHSGEPKDPHDLMTTIHRSTGTISVISGLDREKVPYTLTIOA 300

Qy 301 TMDGDSSTTTAVAVVEILDANDNAPMDFPKYEAHVPENAVGEVORLTVTDLDAVNSP 360
Db 301 TMDGDSSTTTAVAVVEILDANDNAPMDFPKYEAHVPENAVGEVORLTVTDLDAVNSP 360

Qy 361 AWRATYIMGGDDGDHFTTIHPESNQILITTRKGLDPEAKNQHTLYVEVNEAPFVLK 420
Db 361 AWRATYIMGGDDGDHFTTIHPESNQILITTRKGLDPEAKNQHTLYVEVNEAPFVLK 420

Qy 421 PTSTATIIVHVEDVNEAPVFPVPSKVVEQEGIPTEGVCVYTAEDPKENOKISYILR 480
Db 421 PTSTATIIVHVEDVNEAPVFPVPSKVVEQEGIPTEGVCVYTAEDPKENOKISYILR 480

Qy 481 DPAGWLAMPDPSGQVTAAGTLDREDEQFVRNNIYEVWVLAMDNQSPPTTGTGLTLTLLID 540
Db 481 DPAGWLAMPDPSGQVTAAGTLDREDEQFVRNNIYEVWVLAMDNQSPPTTGTGLTLTLLID 540

Qy 541 VNDHGPVPEPRQITICNQSPVRHVLNITDKDLSPTSPFQALTDSDIYWTAEVNEEGD 600
Db 541 VNDHGPVPEPRQITICNQSPVRHVLNITDKDLSPTSPFQALTDSDIYWTAEVNEEGD 600

Qy 601 TVVLSLKKFLKQDVTYDVHLSLSDHGKKEQLTVIRATVCDCHGHVETCPGPKGGFLLPVL 660
Db 601 TVVLSLKKFLKQDVTYDVHLSLSDHGKKEQLTVIRATVCDCHGHVETCPGPKGGFLLPVL 660

Qy 601 TVVLSLKKFLKQDVTYDVHLSLSDHGKKEQLTVIRATVCDCHGHVETCPGPKGGFLLPVL 660
Db 601 TVVLSLKKFLKQDVTYDVHLSLSDHGKKEQLTVIRATVCDCHGHVETCPGPKGGFLLPVL 660

Qy 661 GAVLALLFLLVLLVLLVLRKKIKKEPILLPEDDTRDNVFFYGGEGGEEDQDITQLHR 720
Db 661 GAVLALLFLLVLLVLLVLRKKIKKEPILLPEDDTRDNVFFYGGEGGEEDQDITQLHR 720

Qy 721 GLEARPEVLRNDVAPTIIPPMVRPRANDEIGNFIENLKAANTDPTAPPYDTLLVF 780
Db 721 GLEARPEVLRNDVAPTIIPPMVRPRANDEIGNFIENLKAANTDPTAPPYDTLLVF 780

Qy 781 DYEGSGSDAASLSSTSSASQDQDYDYLNEWGSRRFKKLADMYGGGDD 829
Db 781 DYEGSGSDAASLSSTSSASQDQDYDYLNEWGSRRFKKLADMYGGGDD 829

RESULT 4
US-10-058-270A-126
; Sequence 126, Application US/10058270A
; Publication No. US20040029114A1
; GENERAL INFORMATION:
; APPLICANT: Mack, David H.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Afar, Daniel
```

```

QY 661 GAVLALLFLLVLLVLRKKKIKIPEPLLLPDDTRDNVFFYGGGGGGEDDYDITQIHR 720
DB 661 GAVLALLFLLVLLVLRKKKIKIPEPLLLPDDTRDNVFFYGGGGGGEDDYDITQIHR 720
QY 721 GLEARPEVVLNDVAPTIPTPMYRPRPANDEIGNFIENLKAANTDPTAPPYDITLLVF 780
DB 721 GLEARPEVVLNDVAPTIPTPMYRPRPANDEIGNFIENLKAANTDPTAPPYDITLLVF 780
QY 781 DYEGSGDAASLSLTSSASDQDDYDYLNWGSRFKKLADMYCGGDEDD 829
DB 781 DYEGSGDAASLSLTSSASDQDDYDYLNWGSRFKKLADMYCGGDEDD 829

RESULT 5
US-10-174-677-21
; Sequence 21, Application US/10174677
; Publication No. US20030190704A1
; GENERAL INFORMATION:
; APPLICANT: Xie, Ting
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ANCHORING STEM CELLS IN A MICROENVIR
; FILE REFERENCE: 40716(IP-012)
; CURRENT APPLICATION NUMBER: US/10/174,677
; PRIOR FILING DATE: 2002-06-19
; NUMBER OF SEQ ID NOS: 117
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-174-677-21

Query Match 100.0%; Score 4369; DB 14; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPRGPLASLLLLQVCWLQCAASEPCRAVFEAEVTLKAGABQEPQALGKVFVWGCPG 60
DB 1 MGLPRGPLASLLLLQVCWLQCAASEPCRAVFEAEVTLKAGABQEPQALGKVFVWGCPG 60
QY 61 QEPALFSTDNDDFTVRNGETVOERRSLKERNPLKIFPSKILRRHKRDWVAPISVPENG 120
DB 61 QEPALFSTDNDDFTVRNGETVOERRSLKERNPLKIFPSKILRRHKRDWVAPISVPENG 120
QY 121 KGPPFQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWLKLNKPLDREBIAK 180
DB 121 KGPPFQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWLKLNKPLDREBIAK 180
QY 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTPRGSVLEGVLPGTSMQVAT 240
DB 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTPRGSVLEGVLPGTSMQVAT 240
QY 241 DEDDAIYTYNGVWVAYSIIHSQEPKDPHLMFTIHRSTGTISVISGLDREKVPYTLTIOA 300
DB 241 DEDDAIYTYNGVWVAYSIIHSQEPKDPHLMFTIHRSTGTISVISGLDREKVPYTLTIOA 300
QY 301 TDMGDGSGTTTAVAVVILNDANAPMPDPQKYEAHVPENAVGHVORLTVTDLDPNSP 360
DB 301 TDMGDGSGTTTAVAVVILNDANAPMPDPQKYEAHVPENAVGHVORLTVTDLDPNSP 360
QY 361 AWRATYILMGDDGDHDTITTHPESNGQILTRKGLDFEAKNQHTLVVEVNEAPFVKL 420
DB 361 AWRATYILMGDDGDHDTITTHPESNGQILTRKGLDFEAKNQHTLVVEVNEAPFVKL 420
QY 421 PTSTATTIVHVEDVNEAPFVPPSKVVEQEGIPTGEPVCVYTAEDPDKENQKISYRILR 480
DB 421 PTSTATTIVHVEDVNEAPFVPPSKVVEQEGIPTGEPVCVYTAEDPDKENQKISYRILR 480
QY 481 DPAGWLAMPDPSGGQVAVGLDREDEQFVRNNIYEVNVLANDNGSPPTTGTLTLTLD 540
DB 481 DPAGWLAMPDPSGGQVAVGLDREDEQFVRNNIYEVNVLANDNGSPPTTGTLTLTLD 540
QY 541 VNDHGVPPEPRQITICNQSFPVRHVLNITDKDLSPTSFPQALTDSDIYWTAEVNEEGD 600

```

```

DB 541 VNDHGVPPEPRQITICNQSFPVRHVLNITDKDLSPTSFPQALTDSDIYWTAEVNEEGD 600
QY 601 TVVLSLKKFLKQDITYDVLHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFLLPVL 660
DB 601 TVVLSLKKFLKQDITYDVLHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFLLPVL 660
QY 661 GAVLALLFLLVLLVLRKKKIKIPEPLLLPDDTRDNVFFYGGGGGGEDDYDITQIHR 720
DB 661 GAVLALLFLLVLLVLRKKKIKIPEPLLLPDDTRDNVFFYGGGGGGEDDYDITQIHR 720
QY 721 GLEARPEVVLNDVAPTIPTPMYRPRPANDEIGNFIENLKAANTDPTAPPYDITLLVF 780
DB 721 GLEARPEVVLNDVAPTIPTPMYRPRPANDEIGNFIENLKAANTDPTAPPYDITLLVF 780
QY 781 DYEGSGDAASLSLTSSASDQDDYDYLNWGSRFKKLADMYCGGDEDD 829
DB 781 DYEGSGDAASLSLTSSASDQDDYDYLNWGSRFKKLADMYCGGDEDD 829

RESULT 6
US-10-158-123-1
; Sequence 1, Application US/10158123
; Publication No. US20030194406A1
; GENERAL INFORMATION:
; APPLICANT: Reinhard, Christoph
; APPLICANT: Klinger, Julie
; APPLICANT: Jefferson, Ann
; APPLICANT: Escobedo, Jaime
; APPLICANT: Randazzo, Filippo
; APPLICANT: Winter, Jill
; APPLICANT: Goodson, Robert
; APPLICANT: Qi, Weimin
; TITLE OF INVENTION: P-Cadherin as a Target for Anti-Cancer
; FILE REFERENCE: 35784/258994
; CURRENT APPLICATION NUMBER: US/10/158,123
; CURRENT FILING DATE: 2003-02-07
; PRIOR APPLICATION NUMBER: 60/294,225
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-158-123-1

Query Match 100.0%; Score 4369; DB 14; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPRGPLASLLLLQVCWLQCAASEPCRAVFEAEVTLKAGABQEPQALGKVFVWGCPG 60
DB 1 MGLPRGPLASLLLLQVCWLQCAASEPCRAVFEAEVTLKAGABQEPQALGKVFVWGCPG 60
QY 61 QEPALFSTDNDDFTVRNGETVOERRSLKERNPLKIFPSKILRRHKRDWVAPISVPENG 120
DB 61 QEPALFSTDNDDFTVRNGETVOERRSLKERNPLKIFPSKILRRHKRDWVAPISVPENG 120
QY 121 KGPPFQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWLKLNKPLDREBIAK 180
DB 121 KGPPFQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWLKLNKPLDREBIAK 180
QY 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTPRGSVLEGVLPGTSMQVAT 240
DB 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTPRGSVLEGVLPGTSMQVAT 240
QY 241 DEDDAIYTYNGVWVAYSIIHSQEPKDPHLMFTIHRSTGTISVISGLDREKVPYTLTIOA 300
DB 241 DEDDAIYTYNGVWVAYSIIHSQEPKDPHLMFTIHRSTGTISVISGLDREKVPYTLTIOA 300
QY 301 TDMGDGSGTTTAVAVVILNDANAPMPDPQKYEAHVPENAVGHVORLTVTDLDPNSP 360

```

Db 301 TDMGDSGTTTAVAVVEILDANDNAPMFDQKYEAHPENAVGHEVQELVTDLADPNP 360
Qy 361 AWRATYILMGDDGDHFTITTHPESNOGILTRKGLDPEAKNQHTLYVEVNEAPFVLKL 420
Db 361 AWRATYILMGDDGDHFTITTHPESNOGILTRKGLDPEAKNQHTLYVEVNEAPFVLKL 420
Qy 421 PTSTATIVHVEDVNEAPVFPVPSKVVEVQEGPTGEPVCVYTAEDDPKENOKISYRILR 480
Db 421 PTSTATIVHVEDVNEAPVFPVPSKVVEVQEGPTGEPVCVYTAEDDPKENOKISYRILR 480
Qy 481 DPAGWLAMPDPSGOVTAVGTILDRDEQFVRNNIYEVWVLMNDNGSPPTTGTGLLLTLID 540
Db 481 DPAGWLAMPDPSGOVTAVGTILDRDEQFVRNNIYEVWVLMNDNGSPPTTGTGLLLTLID 540
Qy 541 VNDHGPVPEPQITTCNOGSPVRHVLNITDOLSPHTSPFOAQLTDDSDIYWTAEVNEEGD 600
Db 541 VNDHGPVPEPQITTCNOGSPVRHVLNITDOLSPHTSPFOAQLTDDSDIYWTAEVNEEGD 600
Qy 601 TVVLSLKKFLKQDTYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFLLPVL 660
Db 601 TVVLSLKKFLKQDTYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFLLPVL 660
Qy 661 GAVLALLFLLVLLVLLVLRKKRKEPILLPEDDTRDNVYVYGGEGGEDDQDYDTQLHR 720
Db 661 GAVLALLFLLVLLVLLVLRKKRKEPILLPEDDTRDNVYVYGGEGGEDDQDYDTQLHR 720
Qy 721 GLEAPPEVLRNDVAPTIIPTPMYRPRANPDEIGNFIIENLKAANTDPTAPPYDTLLVF 780
Db 721 GLEAPPEVLRNDVAPTIIPTPMYRPRANPDEIGNFIIENLKAANTDPTAPPYDTLLVF 780
Qy 781 DYEGSGSDAASLSLTSASDQDDYDYLNEWGSRFKKLADMYGGGEDD 829
Db 781 DYEGSGSDAASLSLTSASDQDDYDYLNEWGSRFKKLADMYGGGEDD 829

RESULT 7

US-10-295-027-783
; Sequence 783, Application US/10295027
; Publication No. US2003023250A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of diagnosis of Cancer, Compositions and
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250

; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 783
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-783

Query Match 100.0%; Score 4369; DB 15; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGLPRGSLASILLQVCMLOCAASPCRAVREAVTLEAGCAEOPGALCKVPMGCPG 60
Db 1 MGLPRGSLASILLQVCMLOCAASPCRAVREAVTLEAGCAEOPGALCKVPMGCPG 60
Qy 61 QBPALFSTDNDDFTVRNGETVQERSLKERPLKIFPSKRILRRHKRDVWAPIVSPENG 120
Db 61 QBPALFSTDNDDFTVRNGETVQERSLKERPLKIFPSKRILRRHKRDVWAPIVSPENG 120
Qy 121 KGPFFQRLNQLKSNKDRDTKIFYSITGCAASPPGCVFAVEKETGWLILNKLDRDEETAK 180
Db 121 KGPFFQRLNQLKSNKDRDTKIFYSITGCAASPPGCVFAVEKETGWLILNKLDRDEETAK 180
Qy 181 YELFGHAYSENGASVEDPMNISIIIVTDQNDHKPKTQDTERGSLVLEGLVPGTSVMQVAT 240
Db 181 YELFGHAYSENGASVEDPMNISIIIVTDQNDHKPKTQDTERGSLVLEGLVPGTSVMQVAT 240
Qy 241 DEDDAIYTVNGVAVYSIHSQEPKDPHLMFTIHRSTGTISVSISSGLDREKVEYTLTQA 300
Db 241 DEDDAIYTVNGVAVYSIHSQEPKDPHLMFTIHRSTGTISVSISSGLDREKVEYTLTQA 300
Qy 301 TDMGDSGTTTAVAVVEILDANDNAPMFDQKYEAHPENAVGHEVQELVTDLADPNP 360
Db 301 TDMGDSGTTTAVAVVEILDANDNAPMFDQKYEAHPENAVGHEVQELVTDLADPNP 360
Qy 361 AWRATYILMGDDGDHFTITTHPESNOGILTRKGLDPEAKNQHTLYVEVNEAPFVLKL 420
Db 361 AWRATYILMGDDGDHFTITTHPESNOGILTRKGLDPEAKNQHTLYVEVNEAPFVLKL 420
Qy 421 PTSTATIVHVEDVNEAPVFPVPSKVVEVQEGPTGEPVCVYTAEDDPKENOKISYRILR 480
Db 421 PTSTATIVHVEDVNEAPVFPVPSKVVEVQEGPTGEPVCVYTAEDDPKENOKISYRILR 480
Qy 481 DPAGWLAMPDPSGOVTAVGTILDRDEQFVRNNIYEVWVLMNDNGSPPTTGTGLLLTLID 540
Db 481 DPAGWLAMPDPSGOVTAVGTILDRDEQFVRNNIYEVWVLMNDNGSPPTTGTGLLLTLID 540
Qy 541 VNDHGPVPEPQITTCNOGSPVRHVLNITDOLSPHTSPFOAQLTDDSDIYWTAEVNEEGD 600
Db 541 VNDHGPVPEPQITTCNOGSPVRHVLNITDOLSPHTSPFOAQLTDDSDIYWTAEVNEEGD 600
Qy 601 TVVLSLKKFLKQDTYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFLLPVL 660
Db 601 TVVLSLKKFLKQDTYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFLLPVL 660
Qy 661 GAVLALLFLLVLLVLLVLRKKRKEPILLPEDDTRDNVYVYGGEGGEDDQDYDTQLHR 720
Db 661 GAVLALLFLLVLLVLLVLRKKRKEPILLPEDDTRDNVYVYGGEGGEDDQDYDTQLHR 720
Qy 721 GLEAPPEVLRNDVAPTIIPTPMYRPRANPDEIGNFIIENLKAANTDPTAPPYDTLLVF 780
Db 721 GLEAPPEVLRNDVAPTIIPTPMYRPRANPDEIGNFIIENLKAANTDPTAPPYDTLLVF 780
Qy 781 DYEGSGSDAASLSLTSASDQDDYDYLNEWGSRFKKLADMYGGGEDD 829
Db 781 DYEGSGSDAASLSLTSASDQDDYDYLNEWGSRFKKLADMYGGGEDD 829

RESULT 8
 US-10-295-027-800
 ; Sequence 800, Application US/10295027
 ; Publication No. US20030232350A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Afar, Daniel
 ; APPLICANT: Aziz, Natasha
 ; APPLICANT: Ginsberg, Wendy M.
 ; APPLICANT: Gish, Kurt C.
 ; APPLICANT: Glynn, Richard
 ; APPLICANT: Hevezi, Peter A.
 ; APPLICANT: Mack, David H.
 ; APPLICANT: Murray, Richard
 ; APPLICANT: Watson, Susan R.
 ; APPLICANT: Eos Biotechnology, Inc.
 ; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
 ; FILE OF INVENTION: Methods of Screening for Modulators of Cancer
 ; FILE REFERENCE: 018501-012500US
 ; CURRENT APPLICATION NUMBER: US/10/295,027
 ; CURRENT FILING DATE: 2002-11-13
 ; PRIOR APPLICATION NUMBER: US 09/663,733
 ; PRIOR FILING DATE: 2000-09-15
 ; PRIOR APPLICATION NUMBER: US 60/350,666
 ; PRIOR FILING DATE: 2001-11-13
 ; PRIOR APPLICATION NUMBER: US 60/335,394
 ; PRIOR FILING DATE: 2001-11-15
 ; PRIOR APPLICATION NUMBER: US 60/332,464
 ; PRIOR FILING DATE: 2001-11-21
 ; PRIOR APPLICATION NUMBER: US 60/334,393
 ; PRIOR FILING DATE: 2001-11-29
 ; PRIOR APPLICATION NUMBER: US 60/340,376
 ; PRIOR FILING DATE: 2001-12-14
 ; PRIOR APPLICATION NUMBER: US 60/347,211
 ; PRIOR FILING DATE: 2002-01-08
 ; PRIOR APPLICATION NUMBER: US 60/347,349
 ; PRIOR FILING DATE: 2002-01-10
 ; PRIOR APPLICATION NUMBER: US 60/355,250
 ; PRIOR FILING DATE: 2002-02-08
 ; PRIOR APPLICATION NUMBER: US 60/356,714
 ; PRIOR FILING DATE: 2002-02-13
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 1366
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 800
 ; LENGTH: 829
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-295-027-800
 Query Match 100.0%; Score 4369; DB 15; Length 829;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MGLPRGPLASLLQLQVCLWCAASEPCRAVFAEVLTEAGGAPQEPGQALGVFMGCPG 60
 DB 1 MGLPRGPLASLLQLQVCLWCAASEPCRAVFAEVLTEAGGAPQEPGQALGVFMGCPG 60
 QY 61 QEPALFSTNDNDFVRNGETVOERSLKERPLKIFSKRLRHKRDWVAPISVPENG 120
 DB 61 QEPALFSTNDNDFVRNGETVOERSLKERPLKIFSKRLRHKRDWVAPISVPENG 120
 QY 121 KGPPFQRLNQLKSKNRDRTKIFYSTIGPADSPPEGVFAVEKETGWLKLLNKPLDREBIAK 180
 DB 121 KGPPFQRLNQLKSKNRDRTKIFYSTIGPADSPPEGVFAVEKETGWLKLLNKPLDREBIAK 180
 QY 191 YELFGHVAENGASVEDPMNIIIVTQNDHKPKFTQDTRGSLVGLPGTSVMQVAT 240
 DB 191 YELFGHVAENGASVEDPMNIIIVTQNDHKPKFTQDTRGSLVGLPGTSVMQVAT 240
 QY 241 DEDDAIYNGVWYSHSQBPDPKPHDLMFTIHRSTGTISVSSGLDREKVPYTLTIOA 300
 DB 241 DEDDAIYNGVWYSHSQBPDPKPHDLMFTIHRSTGTISVSSGLDREKVPYTLTIOA 300

QY 301 TDMGDGGSTTTAVAVVEILDANDNAPMEDPQKYEAHVPENAVGHVQRLTVDLDAPNSP 360
 DB 301 TDMGDGGSTTTAVAVVEILDANDNAPMEDPQKYEAHVPENAVGHVQRLTVDLDAPNSP 360
 QY 361 AWRATYLINGDDGDHFTITTHPESNQILTRKGLDPEAKNQHTLYVEVNEAPFVLKL 420
 DB 361 AWRATYLINGDDGDHFTITTHPESNQILTRKGLDPEAKNQHTLYVEVNEAPFVLKL 420
 QY 421 PTSTATIVVHVEDVNEAPVFPVPPSKVVEQIGPTGEPVCVYTAEDDPKQKISYRILR 480
 DB 421 PTSTATIVVHVEDVNEAPVFPVPPSKVVEQIGPTGEPVCVYTAEDDPKQKISYRILR 480
 QY 481 DPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVWVLAMDNQSPPTGTGTLTLLID 540
 DB 481 DPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVWVLAMDNQSPPTGTGTLTLLID 540
 QY 541 VNDHGPVPEPROITICNOSPVRHVLNITDKLSPTSPQALTDSDSDIYWTAEVNEEGD 600
 DB 541 VNDHGPVPEPROITICNOSPVRHVLNITDKLSPTSPQALTDSDSDIYWTAEVNEEGD 600
 QY 601 TVVLSLKXFLKQDQTYDVHLSLSDHGKQELTVIRATVCDCHGVETCPGPKGFFILPVL 660
 DB 601 TVVLSLKXFLKQDQTYDVHLSLSDHGKQELTVIRATVCDCHGVETCPGPKGFFILPVL 660
 QY 661 GAVTALLFLLVLLVLRKKIKIPELLPEDDTRDNVYVYGGSGGSEEDDQDITOLHR 720
 DB 661 GAVTALLFLLVLLVLRKKIKIPELLPEDDTRDNVYVYGGSGGSEEDDQDITOLHR 720
 QY 721 GLEARPEVVLNRDVAPTIPTPMVAPRANPDEIGNFIENLKAANTDFTAPPYDTLLVF 780
 DB 721 GLEARPEVVLNRDVAPTIPTPMVAPRANPDEIGNFIENLKAANTDFTAPPYDTLLVF 780
 QY 781 DYEAGSAAASLSLTSASDQDQDYVNLNWSGRFCKLADMYGGEDD 829
 DB 781 DYEAGSAAASLSLTSASDQDQDYVNLNWSGRFCKLADMYGGEDD 829
 RESULT 9
 US-10-295-027-863
 ; Sequence 863, Application US/10295027
 ; Publication No. US20030232350A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Afar, Daniel
 ; APPLICANT: Aziz, Natasha
 ; APPLICANT: Ginsberg, Wendy M.
 ; APPLICANT: Gish, Kurt C.
 ; APPLICANT: Glynn, Richard
 ; APPLICANT: Hevezi, Peter A.
 ; APPLICANT: Mack, David H.
 ; APPLICANT: Murray, Richard
 ; APPLICANT: Watson, Susan R.
 ; APPLICANT: Eos Biotechnology, Inc.
 ; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
 ; FILE OF INVENTION: Methods of Screening for Modulators of Cancer
 ; FILE REFERENCE: 018501-012500US
 ; CURRENT APPLICATION NUMBER: US/10/295,027
 ; CURRENT FILING DATE: 2002-11-13
 ; PRIOR APPLICATION NUMBER: US 09/663,733
 ; PRIOR FILING DATE: 2000-09-15
 ; PRIOR APPLICATION NUMBER: US 60/350,666
 ; PRIOR FILING DATE: 2001-11-13
 ; PRIOR APPLICATION NUMBER: US 60/335,394
 ; PRIOR FILING DATE: 2001-11-15
 ; PRIOR APPLICATION NUMBER: US 60/332,464
 ; PRIOR FILING DATE: 2001-11-21
 ; PRIOR APPLICATION NUMBER: US 60/334,393
 ; PRIOR FILING DATE: 2001-11-29
 ; PRIOR APPLICATION NUMBER: US 60/340,376
 ; PRIOR FILING DATE: 2001-12-14
 ; PRIOR APPLICATION NUMBER: US 60/347,211
 ; PRIOR FILING DATE: 2002-01-08
 ; PRIOR APPLICATION NUMBER: US 60/347,349
 ; PRIOR FILING DATE: 2002-01-10

; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 863
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-863

Query Match
Best Local Similarity 100.0%; Score 4369; DB 15; Length 829;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGLPRGLASLLLLQVCMWLCQAASEPCRAVFREAEVTLAAGGAEQPGQALGVFMGCGP 60
Db 1 MGLPRGLASLLLLQVCMWLCQAASEPCRAVFREAEVTLAAGGAEQPGQALGVFMGCGP 60

Qy 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKILRRHKRDWVAPISVPENG 120
Db 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKILRRHKRDWVAPISVPENG 120

Qy 121 KGPPFQRLNQLKSKNRDRTKIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDREEIAK 180
Db 121 KGPPFQRLNQLKSKNRDRTKIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDREEIAK 180

Qy 121 KGPPFQRLNQLKSKNRDRTKIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDREEIAK 180
Db 121 KGPPFQRLNQLKSKNRDRTKIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDREEIAK 180

Qy 181 YELFGHVAIVNGVAVSIHSEKQDHLMTIHRSTGTISVISGLDREKVPYTLTQA 300
Db 181 YELFGHVAIVNGVAVSIHSEKQDHLMTIHRSTGTISVISGLDREKVPYTLTQA 300

Qy 181 YELFGHVAIVNGVAVSIHSEKQDHLMTIHRSTGTISVISGLDREKVPYTLTQA 300
Db 181 YELFGHVAIVNGVAVSIHSEKQDHLMTIHRSTGTISVISGLDREKVPYTLTQA 300

RESULT 10
US-10-295-027-896
; Sequence 896, Application US/10295027
; Publication NO. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Nataasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; FILE REFERENCE: 018501-01250005
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 896
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-896

Query Match
Best Local Similarity 100.0%; Score 4369; DB 15; Length 829;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGLPRGLASLLLLQVCMWLCQAASEPCRAVFREAEVTLAAGGAEQPGQALGVFMGCGP 60
Db 1 MGLPRGLASLLLLQVCMWLCQAASEPCRAVFREAEVTLAAGGAEQPGQALGVFMGCGP 60

Qy 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKILRRHKRDWVAPISVPENG 120
Db 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKILRRHKRDWVAPISVPENG 120

Qy 121 KGPPFQRLNQLKSKNRDRTKIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDREEIAK 180
Db 121 KGPPFQRLNQLKSKNRDRTKIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDREEIAK 180

Qy 181 YELFGHVAIVNGVAVSIHSEKQDHLMTIHRSTGTISVISGLDREKVPYTLTQA 300
Db 181 YELFGHVAIVNGVAVSIHSEKQDHLMTIHRSTGTISVISGLDREKVPYTLTQA 300

```
QY 301 TMDGSGSTTTAVAVVEILDANDNAPMFDPOKYEAHVPENAVGEVORLTVTDLDAENSP 360
DB 301 TMDGSGSTTTAVAVVEILDANDNAPMFDPOKYEAHVPENAVGEVORLTVTDLDAENSP 360
QY 361 AWRATYLMGDDGDDGHFTITTHPESNOGILTRKGLDPEAKNQHTLYVEVNEAPFVLK 420
DB 361 AWRATYLMGDDGDDGHFTITTHPESNOGILTRKGLDPEAKNQHTLYVEVNEAPFVLK 420
QY 421 PTSTATIIVHVEDVNEAPVFPVPPSKVVEVQEGITPGEVPCVYTAEDDPKENOKISYRILR 480
DB 421 PTSTATIIVHVEDVNEAPVFPVPPSKVVEVQEGITPGEVPCVYTAEDDPKENOKISYRILR 480
QY 481 DPAGWLAMPDSGGQVAVGTLDREDEQFVRNNIYEVVWVLANMDSPPTTGTGTLTLLID 540
DB 481 DPAGWLAMPDSGGQVAVGTLDREDEQFVRNNIYEVVWVLANMDSPPTTGTGTLTLLID 540
QY 541 VNDHGPVPEPRQITICNOSPVRHVNLITDKLSHTSPFOAQLTDDSDIYWTAEVNEEGD 600
DB 541 VNDHGPVPEPRQITICNOSPVRHVNLITDKLSHTSPFOAQLTDDSDIYWTAEVNEEGD 600
QY 601 TVVLSLKKFLKQDTYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFILLPV 660
DB 601 TVVLSLKKFLKQDTYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFILLPV 660
QY 661 GAVLALLFLLVLLVLLVLRKKKIKIPELLLPEDDTRDNVFFYGGEGGEEQDYDITQLHR 720
DB 661 GAVLALLFLLVLLVLLVLRKKKIKIPELLLPEDDTRDNVFFYGGEGGEEQDYDITQLHR 720
QY 721 GLEAPPEVLRNDVAPTIIPTPMYRPRPANPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
DB 721 GLEAPPEVLRNDVAPTIIPTPMYRPRPANPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
QY 781 DYESGSDAASLSLTSSASDQDYDYLNEWGSFRFKKLADMYGGEDD 829
DB 781 DYESGSDAASLSLTSSASDQDYDYLNEWGSFRFKKLADMYGGEDD 829

RESULT 11
US-10-229-345-16
; Sequence 16, Application US/10229345
; Publication No. US20040038220A1
; GENERAL INFORMATION:
; APPLICANT: MARKOWITZ, Sanford D.
; TITLE OF INVENTION: METHODS FOR CATEGORIZING PATIENTS
; FILE REFERENCE: CWRU-P01-003
; CURRENT APPLICATION NUMBER: US/10/229,345
; PRIOR FILING DATE: 2002-08-26
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-229-345-16

Query Match 99.8%; Score 4361; DB 12; Length 829;
Best Local Similarity 99.9%; Pred No. 0;
Matches 828; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGLPRGPPLASLLLLQVCWLQCAASEPCRAVFEAVTLEAGGAEQEPGALGVFMGCPG 60
DB 1 MGLPRGPPLASLLLLQVCWLQCAASEPCRAVFEAVTLEAGGAEQEPGALGVFMGCPG 60
QY 61 QEPALFSTNDNDDFTVRNGETVQERSLSKERNPLKIFPSKRILSRHKEDWVAVPISVPENG 120
DB 61 QEPALFSTNDNDDFTVRNGETVQERSLSKERNPLKIFPSKRILSRHKEDWVAVPISVPENG 120
QY 121 KGPPFPQRLNQLSKNRDRTKIFYITGFGADSPPEGVFAVEKETGWLKLLNKLPLDREIAK 180
DB 121 KGPPFPQRLNQLSKNRDRTKIFYITGFGADSPPEGVFAVEKETGWLKLLNKLPLDREIAK 180
QY 181 YELFGHVAENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSLVLEGVLPGTSVMQVTAT 240
```

```
DB 181 YELFGHVAENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSLVLEGVLPGTSVMQVTAT 240
QY 241 DEDDAIYTYNGVWVAYSIIHQBPCKDPHDLMFTHIRSTGTISVSSGLDREKPEVTLTIOA 300
DB 241 DEDDAIYTYNGVWVAYSIIHQBPCKDPHDLMFTHIRSTGTISVSSGLDREKPEVTLTIOA 300
QY 301 TMDGSGSTTTAVAVVEILDANDNAPMFDPOKYEAHVPENAVGEVORLTVTDLDAENSP 360
DB 301 TMDGSGSTTTAVAVVEILDANDNAPMFDPOKYEAHVPENAVGEVORLTVTDLDAENSP 360
QY 361 AWRATYLMGDDGDDGHFTITTHPESNOGILTRKGLDPEAKNQHTLYVEVNEAPFVLK 420
DB 361 AWRATYLMGDDGDDGHFTITTHPESNOGILTRKGLDPEAKNQHTLYVEVNEAPFVLK 420
QY 421 PTSTATIIVHVEDVNEAPVFPVPPSKVVEVQEGITPGEVPCVYTAEDDPKENOKISYRILR 480
DB 421 PTSTATIIVHVEDVNEAPVFPVPPSKVVEVQEGITPGEVPCVYTAEDDPKENOKISYRILR 480
QY 481 DPAGWLAMPDSGGQVAVGTLDREDEQFVRNNIYEVVWVLANMDSPPTTGTGTLTLLID 540
DB 481 DPAGWLAMPDSGGQVAVGTLDREDEQFVRNNIYEVVWVLANMDSPPTTGTGTLTLLID 540
QY 541 VNDHGPVPEPRQITICNOSPVRHVNLITDKLSHTSPFOAQLTDDSDIYWTAEVNEEGD 600
DB 541 VNDHGPVPEPRQITICNOSPVRHVNLITDKLSHTSPFOAQLTDDSDIYWTAEVNEEGD 600
QY 601 TVVLSLKKFLKQDTYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFILLPV 660
DB 601 TVVLSLKKFLKQDTYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFILLPV 660
QY 661 GAVLALLFLLVLLVLLVLRKKKIKIPELLLPEDDTRDNVFFYGGEGGEEQDYDITQLHR 720
DB 661 GAVLALLFLLVLLVLLVLRKKKIKIPELLLPEDDTRDNVFFYGGEGGEEQDYDITQLHR 720
QY 721 GLEAPPEVLRNDVAPTIIPTPMYRPRPANPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
DB 721 GLEAPPEVLRNDVAPTIIPTPMYRPRPANPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
QY 781 DYESGSDAASLSLTSSASDQDYDYLNEWGSFRFKKLADMYGGEDD 829
DB 781 DYESGSDAASLSLTSSASDQDYDYLNEWGSFRFKKLADMYGGEDD 829

RESULT 12
US-10-274-177-16
; Sequence 16, Application US/10274177
; Publication No. US20040038225A1
; GENERAL INFORMATION:
; APPLICANT: MARKOWITZ, Sanford D.
; TITLE OF INVENTION: METHODS FOR CATEGORIZING PATIENTS
; FILE REFERENCE: CWRU-P01-003
; CURRENT APPLICATION NUMBER: US/10/274,177
; PRIOR FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: US/10/229,345
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-274-177-16

Query Match 99.8%; Score 4361; DB 12; Length 829;
Best Local Similarity 99.9%; Pred No. 0;
Matches 828; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGLPRGPPLASLLLLQVCWLQCAASEPCRAVFEAVTLEAGGAEQEPGALGVFMGCPG 60
DB 1 MGLPRGPPLASLLLLQVCWLQCAASEPCRAVFEAVTLEAGGAEQEPGALGVFMGCPG 60
QY 61 QEPALFSTNDNDDFTVRNGETVQERSLSKERNPLKIFPSKRILSRHKEDWVAVPISVPENG 120
```



```

RESULT 14
US-10-295-027-338
; Sequence 338, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 338
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-338

Query Match      99.8%; Score 4361; DB 15; Length 829;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 828; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MGLPRGLASLLILQVLCWCAASEPCRAVREAEVTLKAGGAEQPGQALGKVFVGCPC 60
Db 1 MGLPRGLASLLILQVLCWCAASEPCRAVREAEVTLKAGGAEQPGQALGKVFVGCPC 60

Qy 61 QEPALFSTDDFTVRNGETVQERRSLKERNPLKIPPSKILRRHRKRDWVAPISVPENG 120
Db 61 QEPALFSTDDFTVRNGETVQERRSLKERNPLKIPPSKILRRHRKRDWVAPISVPENG 120

Qy 121 KGPPQSLNOLKSKNRDRTKIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDREIAK 180
Db 121 KGPPQSLNOLKSKNRDRTKIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDREIAK 180

Qy 181 YELFGHAVSENGASVEDPMNISITVDQNDHKPKFTQDTRFGSVLGGVPGTSMQVAT 240
Db 181 YELFGHAVSENGASVEDPMNISITVDQNDHKPKFTQDTRFGSVLGGVPGTSMQVAT 240

Qy 241 DEDDAITYNGVAVSYTHSQEPKDPHDLMTFTHRSCTTSSVSSGLDREKVPYTLTICA 300
Db 241 DEDDAITYNGVAVSYTHSQEPKDPHDLMTFTHRSCTTSSVSSGLDREKVPYTLTICA 300

```

```

Qy 301 TMDGDSGTTTAVAVVEILDANDNAPMEDPKQYEAHVPENAVGHVORLTVTDLDAENSP 360
Db 301 TMDGDSGTTTAVAVVEILDANDNAPMEDPKQYEAHVPENAVGHVORLTVTDLDAENSP 360
Qy 361 AWRATYILMGDDGDHFTITTHPESNQILTRKGLDPEAKNQHTLYVEVTNEAPFVYKL 420
Db 361 AWRATYILMGDDGDHFTITTHPESNQILTRKGLDPEAKNQHTLYVEVTNEAPFVYKL 420
Qy 421 PTSTATIVVHVEDVNEAPVFPVPSKVVEVQEGITGEPVCVYTAEDDPKQKISYILR 480
Db 421 PTSTATIVVHVEDVNEAPVFPVPSKVVEVQEGITGEPVCVYTAEDDPKQKISYILR 480
Qy 481 DPAGWLAMPDPSGOVTVAGTLDRDEQFVRNNIYEVNVLAMNDGSPPTTGTGTLILLID 540
Db 481 DPAGWLAMPDPSGOVTVAGTLDRDEQFVRNNIYEVNVLAMNDGSPPTTGTGTLILLID 540
Qy 541 VNDHGPVPEPQITTCNQSPVRHVLNITDKLSPTSPFQALQTTDSDIYWTAEVNEEG 600
Db 541 VNDHGPVPEPQITTCNQSPVRHVLNITDKLSPTSPFQALQTTDSDIYWTAEVNEEG 600
Qy 601 TVVLSLKKFLKQDVTYDVHLSLSHDHGNKEQLTVIRATVCDCHGHVETCPGWKGGFILPVL 660
Db 601 TVVLSLKKFLKQDVTYDVHLSLSHDHGNKEQLTVIRATVCDCHGHVETCPGWKGGFILPVL 660
Qy 661 GAVLALLFLLVLLVLRKKGKIKEPILLPEDDTRDNVFFYGGEGGEEDQDYDITQLHR 720
Db 661 GAVLALLFLLVLLVLRKKGKIKEPILLPEDDTRDNVFFYGGEGGEEDQDYDITQLHR 720
Qy 721 GLEARPEVVLNDVAPTITPTMYRPRPANDPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
Db 721 GLEARPEVVLNDVAPTITPTMYRPRPANDPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
Qy 781 DYEGSGSDAASLSLTSASDQDQDYDLNENWGSRFKKLADMYGGGDD 829
Db 781 DYEGSGSDAASLSLTSASDQDQDYDLNENWGSRFKKLADMYGGGDD 829

```

```

RESULT 15
US-10-295-027-1246
; Sequence 1246, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10

```

Search completed: September 21, 2004, 22:20:46
Job time : 147 secs

;; PRIOR APPLICATION NUMBER: US 60/355,250
;; PRIOR FILING DATE: 2002-02-08
;; PRIOR APPLICATION NUMBER: US 60/356,714
;; PRIOR FILING DATE: 2002-02-13
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 1386
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 1246
;; LENGTH: 829
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-295-027-1246

Query Match 99.8%; Score 4361; DB 15; Length 829;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 828; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	MGLPRGLASLLLOVCWLCQASEPCRAVFEAEVTLGAGAEQEPGQALGKVFNGCPG	60
Db	1	MGLPRGLASLLLOVCWLCQASEPCRAVFEAEVTLGAGAEQEPGQALGKVFNGCPG	60
QY	61	QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDWVWAPISVPENG	120
Db	61	QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDWVWAPISVPENG	120
QY	121	KGFPORLNQKSNKDRDKIFYSITGPGADSPGEGFAVEKETGWLILNKLPLDRREEIAK	180
Db	121	KGFPORLNQKSNKDRDKIFYSITGPGADSPGEGFAVEKETGWLILNKLPLDRREEIAK	180
QY	181	YELFGHAVSENGASVEDPNMISIIIVTDQNDHKPKFTQDFRGSVLGVLPGTSVMQVAT	240
Db	181	YELFGHAVSENGASVEDPNMISIIIVTDQNDHKPKFTQDFRGSVLGVLPGTSVMQVAT	240
QY	241	DEDAIYTYNGVAYSIHSGEPKDPHDLMTIHRSTGTISVISSSGLDREKVPYTLTIOA	300
Db	241	DEDAIYTYNGVAYSIHSGEPKDPHDLMTIHRSTGTISVISSSGLDREKVPYTLTIOA	300
QY	301	TDMDGSGSTTAVAVVEILDANDNAPMDFPKQYEAHVPENAVGEVQRLTVTDLDAENSP	360
Db	301	TDMDGSGSTTAVAVVEILDANDNAPMDFPKQYEAHVPENAVGEVQRLTVTDLDAENSP	360
QY	361	AWRATVLMGDDGDHFTITTHPESNQGLTTRKGLDFAKNQHTLYVEVTNEAPFVLKL	420
Db	361	AWRATVLMGDDGDHFTITTHPESNQGLTTRKGLDFAKNQHTLYVEVTNEAPFVLKL	420
QY	421	PTSTATIVHVEDVNEAPVFPSPKVVEQEGIPTGECVCTAEDPDKENQKISYRIILR	480
Db	421	PTSTATIVHVEDVNEAPVFPSPKVVEQEGIPTGECVCTAEDPDKENQKISYRIILR	480
QY	481	DPAGWLAMPDSGQVAVGTLDREDEQFVRNNIYEVWVLAMDNQSPPTTGTGLTLTLD	540
Db	481	DPAGWLAMPDSGQVAVGTLDREDEQFVRNNIYEVWVLAMDNQSPPTTGTGLTLTLD	540
QY	541	VNDHGPVPEPRQITICNQSPVRHVLNITDKDLSFHSPPQALTDSDIYWTAEVNEGD	600
Db	541	VNDHGPVPEPRQITICNQSPVRHVLNITDKDLSFHSPPQALTDSDIYWTAEVNEGD	600
QY	601	TVVLSLKKFLKQDYDVHLSLSDHGNKEQLTVIRATVCDCHGVETCPGPKWGGFILPVL	660
Db	601	TVVLSLKKFLKQDYDVHLSLSDHGNKEQLTVIRATVCDCHGVETCPGPKWGGFILPVL	660
QY	661	GAVLALLFLLVLLVLLVLRKKIKPELPLLPEDDTRDNVFFYEGEGGEDQDYITQLHR	720
Db	661	GAVLALLFLLVLLVLLVLRKKIKPELPLLPEDDTRDNVFFYEGEGGEDQDYITQLHR	720
QY	721	GLEARPEVLNDVAPTIIPTMYRPRPANDPEIGNFIIENLKAANTDPTAPPYDILLVF	780
Db	721	GLEARPEVLNDVAPTIIPTMYRPRPANDPEIGNFIIENLKAANTDPTAPPYDILLVF	780
QY	781	DYEGSGSDAASLSLTSASDQDDYDYLNEWGSRFKKLADMYGGGEDD	829
Db	781	DYEGSGSDAASLSLTSASDQDDYDYLNEWGSRFKKLADMYGGGEDD	829

This Page Blank (usp10)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 23, 2004, 03:15:06 ; Search time 5081 Seconds
(without alignments)
4872.220 Million cell updates/sec

Title: US-09-916-849A-1
Perfect score: 829
Sequence: 1 MGLPRGPLASLLLLQVCWLQ.....NEWGSRFKKLADMYGGGDD 829

Scoring table:
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Zgapop 6.0 , Zgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Word size: 1

Total number of hits satisfying chosen parameters: 55025477

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-O=/cpn2.1/uspto.spool.h/US09916849/runat_21092004.145835.7328/app_query.fasta_1.967
-DB=EST -QFMT=fastap -SUFFIX=oligo.rst -MINMATCH=0 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -ENDS=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09916849 @CGN 1.1 3609 @runat_21092004.145835.7328 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_pig:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	659	79.5	3632	11	BC014462 Homo sapi
2	243	29.3	913	13	BQ676796 AGENCOURT
3	238	28.7	918	14	CA454560 AGENCOURT
4	235	28.3	1016	12	BM923882 AGENCOURT
5	228	27.5	875	14	CA489015 AGENCOURT
6	228	27.5	1122	12	BM545820 AGENCOURT
7	227	27.4	769	12	BG403092 AGENCOURT
8	224	27.0	887	14	CA455046 AGENCOURT
9	221	26.7	865	13	BUS43746 AGENCOURT
10	214	25.8	703	12	BM715802 UI-E-EJO-
11	214	25.8	1006	13	BU168948 AGENCOURT
12	213	25.7	745	14	CB240552 UI-CF-FNO
13	211	25.5	880	14	CB92509 AGENCOURT
14	211	25.5	894	14	CA489124 AGENCOURT
15	208	25.1	999	12	BM804940 AGENCOURT
16	204	24.6	836	12	BM045728 AGENCOURT
17	200	24.1	846	13	BQ690885 AGENCOURT
18	197	23.8	892	14	CA489338 AGENCOURT
19	196	23.6	815	14	CB997053 AGENCOURT
20	196	23.6	835	14	CB996355 AGENCOURT
21	196	23.6	949	13	BQ679513 AGENCOURT
22	195	23.5	899	14	CA454360 AGENCOURT
23	180	21.7	854	14	CA488032 AGENCOURT
24	167	20.1	1040	12	BM909817 AGENCOURT
25	166	20.0	778	12	BQ072354 AGENCOURT
26	160	19.3	1001	13	BQ072354 AGENCOURT
27	151	18.2	475	10	BF915077 IL3-UT011
28	151	18.1	873	14	CA453911 AGENCOURT
29	150	18.1	773	10	BE740340 AGENCOURT
30	149	18.0	674	9	AU141604 AGENCOURT
31	149	18.0	902	14	CB994831 AGENCOURT
32	144	17.4	433	12	BI062473 IL3-UT011
33	142	17.1	912	13	BQ928283 AGENCOURT
34	141	17.0	611	10	AW950190 EST382155
35	140	16.9	927	14	CA489357 AGENCOURT
36	135	16.3	901	14	CA454213 AGENCOURT
37	133	16.0	870	10	BE378969 AGENCOURT
38	131	15.8	916	13	BU179183 AGENCOURT
39	126	15.2	848	12	BG675452 602621781
40	122	14.7	888	14	CA487856 AGENCOURT
41	122	14.7	1049	13	BQ936686 AGENCOURT
42	121	14.6	364	12	BG319601 4dh29 Hum
43	118	14.2	761	10	BE741190 601593814
44	117	14.1	790	14	CF593464 AGENCOURT
45	117	14.1	878	14	CA489425 AGENCOURT

ALIGNMENTS

RESULT 1
BC014462
LOCUS
DEFINITION Homo sapiens cadherin 3, type 1, P-cadherin (placental), mRNA (cdna
clone IMAGE:4870356), containing frame-shift errors.
ACCESSION BC014462
VERSION BC014462.1 GI:15680220
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3632)

AUTHORS

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, K.H., Shenmen, C.M., Schuler, G.D.,
Altshuler, S.P., Zebraf, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Carnetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Abramson, R.D., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
McKernan, K.J., Mallek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S.S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzy, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettman, M., Madan, A., Rodrigues, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzyzinski, M.I., Skalska, U., Smalls, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

22388257

12477932

2 (bases 1 to 3632)

Strausberg, R.

Direct Submission

Submitted (17-SEP-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

NIH-MGC project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgabbs-r@mail.nih.gov

Tissue Procurement: ARCC/DCRP/BTP

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Genome Sequence Centre,

BC Cancer Agency, Vancouver, BC, Canada

info@bcsc.bc.ca

Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzyzinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
Schein, Duane Smalls, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranada Tesai, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 34 Row: 1 Column: 14
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 14589890
This clone has the following problem: frame shifted.

FEATURES

source

1. .3632
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4870356"
/tissue_type="skin, melanotic melanoma, high MDR."
/clone_lib="NIH MGC 49"
/lab_host="DH10B-R"
/note="Vector: pOTB7"

ORIGIN

Alignment Scores:

Pred. No.:	0	Length:	3632
Score:	659.00	Matches:	759
Percent Similarity:	99.74%	Conservative:	0
Best Local Similarity:	99.74%	Mismatches:	1
Query Match:	79.49%	Indels:	2
DB:	11	Gaps:	0

US-09-916-849A-1 (1-829) x BC014462 (1-3632)

QY 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuLeuGlnValCysTyrLeuGln 20

DB 388 ATGGGGCTCCCTCGTGACCTCTCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 447

QY 21 CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40

DB 448 TGGCGGGCTCCGAGCGCTCCGCGGGCTCTTTCAGGAGGCTGAAGTGAACCTTGGAGGCG 507

QY 41 GlyGlyAlaGluGlnGluProGlyGlnAlaLeuGlyValPheMetGlyCysProGly 60

DB 508 GGAGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 567

QY 61 GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr 80

DB 568 CAAGAGCCAGCTCTGTTAGCACTGATATGATGACTTCACTGTGCGGAATGCGGAGACA 627

QY 81 ValGlnGluArgSerLeuLysGluArgAsnProLeuLysLysLysLysLysLysLys 100

DB 628 GTCCAGGAAAGAGGTCACCTGAAGGAAAGGAATCCATTGAAGATCTTCCCATCCAAAGT 687

QY 101 IleLeuArgArgHisLysArgAspTyrValAlaProIleSerValProGluAsnGly 120

DB 688 ATCTTAGAGAGACACAGAGAGATGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 747

QY 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140

DB 748 AAGGGTCCCTTCCCGCAGAGACTGAATCAAGTCAAGTCAATTAAGATAGAGACACCAAG 807

QY 141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProGluGlyValPheAlaVal 160

DB 808 ATTTTCTACAGCATCAGGGGGCGGGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 867

QY 161 GluLysGluThrGlyTyrLeuLeuLeuAsnLysProLeuAspArgGluLysLysLys 180

DB 868 GAG 927

QY 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200

DB 928 TATGAGCTCTTTGGCCACGCTGTGTGAGAGATGGTGGTGGTGGTGGTGGTGGTGGTGGT 987

QY 201 IleSerIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220

DB 988 ATCTCCATCATCTGTGACGAGAGAGATGACCAAGCCCAAGTTTACCAGGAGACACCTTC 1047

QY 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240

DB 1048 CGAGGGAGTGTCTTAGAGGGAGTCTTACCAGGAGTCTTGTGTGTGTGTGTGTGTGTGTGT 1107

QY 241 AspGluAspAlaIleTyrThrTyrAsnGlyValAlaValSerIleHisSerGln 260

DB 1108 GATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1167

QY 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280

DB 1168 GAACCAAG 1227

QY 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300

DB 1228 GTATCTCCAGTGGCTGGACCGGAGAAAGTCCCTGAGTACACACTGACCATCAGGAGGC 1287

QY 301 ThrAspMetAspGlyAspGlySerThrThrAlaValAlaValValGluLeuLeuAsp 320

DB 1288 ACAGACATGATGGGAGCGGCTCCACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1347

QY 321 AlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340

DB 1348 GCCAATGACATGCTCCCATGTTTACCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1407

QY 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360

Db 1408 GCAGTGGGCGCATGAGGTGCAGAGGTGACGGTCACTGATCTCGACGCCGCCCACTCACCA 1467
Qy 361 AlaTpaAtaAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
Db 1468 GGTGGCGTGCACCTACCTTATCATATGGCGGTGACGACGGGACCATTTTACCATCAC 1527
Qy 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
Db 1528 ACCACCCCTGAGAGCAACACGAGGCTCTCTGCAACACGAGAGGGTTTGATTTTGGAGCC 1587
Qy 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db 1588 AAAAACCCAGCACACCTGTGACGTGAAGTGACCAAGAGGCCCTTTTGTGTGAAGCTC 1647
Qy 421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
Db 1648 CCAACCTCCACAGCCACCATAGTGTCCACGTGAGGATGTGAATGAGGCACCTGTGTT 1707
Qy 441 ValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
Db 1708 GTCCCAACCTCCAAAGTCTGAGTCCAGGAGGCACTCCACACTGGGAGGCTGTGTGT 1767
Qy 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db 1768 GTCTACACTGCAGAGACCTGCACAGGAGATCAAGATCAGCTACCGCATCTCGAGA 1827
Qy 481 AspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db 1828 GACCCAGCAGGCTGTAGCTAGCCATGACCCAGACAGTGGCAGGTACAGCTGTGGCCACC 1887
Qy 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db 1888 CTCGACCGTGGAGATGAGCGTTTGGAGGACACATCATGAGTATGCTGTGTCGCC 1947
Qy 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuThrLeuIleAsp 540
Db 1948 ATGGCAATGTAAGCCCTCCACCACTGCGCAGGAAACCTTCTGTAACACTGATTGAT 2007
Qy 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 2008 GTCAACAGCATGGCCAGCTCCCTGAGCCCTGCAGATCACCATCTGCAACCAAGCCCT 2067
Qy 561 ValArgHis-ValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
Db 2068 GTGGCGCA-GTGTCTGAACATACGACACAGACCTGTCTCCACACCTCCCTTTCCA 2126
Qy 580 nlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGlyAs 600
Db 2127 GGCCCACTCAGATGACTCAGACATCTACTGCGAGGAGGTCAACAGGAAGGTGA 2186
Qy 600 pThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSe 620
Db 2187 CACAGTGGTCTTGTCTCCCTGAGAAGTCTCTGAGCAGGATACATATGCTGCACCTTTC 2246
Qy 620 rLeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCy 640
Db 2247 TCTGTCTGACCATGGCAACAAAGAGCAGCTGACGGTGATCAGGGCCACTGTGTGGAGCTG 2306
Qy 640 shisGlyHisValGluThrCysProGlyProTrpLysGlyGlyPheIleLeuProValle 660
Db 2307 CCATGGCCATGTGAAACCTGCCCTGGACCCCTGGAAGAGGATTCATCTCCCTCTGTCT 2366
Qy 660 uGlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysLys 680
Db 2367 GGGGGTCTGCTGCTCTCTCTCTCTGCTGCTGCTGCTTGTGTTGGTGAAGAAGAA 2426
Qy 680 sArgLysIleLysGluProLeuLeuLeuProGluLeuAspThrArgAspAsnValPheTy 700
Db 2427 GCGGAAGATCAGAGAGCCCTCTACTCCAGAGATGACACCGGTGACACAGCTCTTCTTA 2486
Qy 700 rTyrGlyGluGluGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisAr 720
Db 2487 CTATGCGGAGAGGGGGTGGGAGAGGACCGAGACTATGACATCACCACCTCCACCG 2546

Qy 720 gGlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePr 740
Db 2547 AGTCTGGAGCGCGGCGGAGGTGTCTCCGCAATGACGTGGCACCACCATCATCCC 2606
Qy 740 oThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleG 760
Db 2607 GACACCATGTACCGTCTCTAGCCAGCAACCCAGATGAATCGCAACTTTTATAATTGA 2666
Qy 760 u 760
Db 2667 G 2667

RESULT 2
BQ676796 913 bp mRNA linear EST 15-JUL-2002
LOCUS AGENCOURT 8184299 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6264044
DEFINITION 5' mRNA Sequence.

ACCESSION BQ676796
VERSION BQ676796.1 GI:21789475
KEYWORDS EST.
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 913)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

Plate: LICM2428 row: 0 column: 21
High quality sequence stop: 635.
Location/Qualifiers

FEATURES
source
1..913
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6264044"
/tissue_type="melanotic melanoma, cell line"
/lab_host="NIH_MGC_112"
/note="Organ: skin; Vector: pOTB7; Site: 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

ORIGIN
Alignment Scores:
Pred. No.: 9,49e-208 Length: 913
Score: 243.00 Matches: 243
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 29.31% Indels: 0
DB: 13 Gaps: 0

US-09-916-849A-1 (1-829) x BQ676796 (1-913)

Qy 275 SerThrGlyThrIleSerValIleSerSerGlyLeuAspArgGluLysValProGluTyr 294
Db 2 AGCACAGGCACCATCAGCGTCATCTCCAGTGGCTGGACCGGAAAAAGTCCCTGAGTAC 61

Qy 295 ThrLeuThrIleGlnAlaThrAspMetAspGlyAspGlySerThrThrThrAlaValAla 314
 Db 62 ACACCTGACCATCCAGGCCACAGACATGGATGGGACGGCTCCACACACCGCAGTGGCA 121
 Qy 315 ValValGluIleLeuAspAlaThrAspMetAspGlyAspGlySerThrThrAlaValAla 334
 Db 122 GTAGTGGAGATCTTGTATGCCAATGACATGCTCCATGTTTGCACCCCGAAGATACGAG 181
 Qy 335 AlaHisValProGluAsnAlaValGlyHisGluValGlnAlaThrValThrAspLeu 354
 Db 182 GCCCATGTGCTGAGAATGAGTGGGCCATGAGTGCAGAGGCTGACGGTCACTGATCTG 241
 Qy 355 AspAlaProAsnSerProAlaThrArgAlaThrTyLeuIleMetClyAspAspGly 374
 Db 242 GAGCGCCCCAACCTACACAGCGTGGCGTGCACCTACTTATCATGGGCGGTGACACGGG 301
 Qy 375 AspHisPheThrIleThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLys 394
 Db 302 GACCAITTTTACCATACCCACCCCTGAGAGCAACAGGGCATCTCTGACACACAGGAG 361
 Qy 395 GlyLeuAspPheGluAlaLysAsnGlnHisThrLeuTyValGluValThrAsnGluAla 414
 Db 362 GGTITGGATTTTGGAGGCCAANAACACAGCACCCCTGTACGTTTGAAGTGCACCAACAGGCC 421
 Qy 415 PropheValLeuLysLeuProThrSerThrAlaThrIleValValHisValGluAspVal 434
 Db 422 CTTTGTGTCTGAAGTCTCCCACTCCAGCCACAGCCACCATAGTGTCTCAGTGGAGATGTG 481
 Qy 435 AsnGluAlaProValPheValProSerLysValValGluValGlnGluGlyIlePro 454
 Db 482 AATGAGGCACCTGTGTTTGTCCACCCCTCCAAAGTCGTTGAGTTCAGAGGGCATCCCC 541
 Qy 455 ThrGlyGluProValCysValTyThrAlaGluAspProAspLysGluAsnGlnLysIle 474
 Db 542 ACTGGGAGCCTGTGTGTCTACATCTGCAGAGAGCCCTGCAGAGAGATCAAAAGATC 601
 Qy 475 SerTyArgIleLeuArgAspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGln 494
 Db 602 AGCTACCGCATCTGAGAGACCCAGCAGCGGTGGTAGCATGGACCCAGACAGTGGCAG 661
 Qy 495 ValThrAlaValGlyThrLeuAspArgGluAspGluGlnPheValArgAsnHisIleTy 514
 Db 662 GTCACAGCTGTGGCACCCCTGCACCGTGGAGTGGAGCAGTGTGTGGAGGAACAATCTAT 721
 Qy 515 GluValMet 517
 Db 722 GAAGTCATG 730

RESULT 3
 CA454560
 LOCUS
 DEFINITION AGENCOURT 10714108 MAPcL Homo sapiens cDNA clone IMAGE:6720636 5',
 mRNA sequence.
 ACCESSION CA454560
 VERSION CA454560.1 GI:24904418
 KEYWORDS EST.
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 918)
 NIH-MGC <http://mgs.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Kristi A. Eglund, Ira Pastan
 CDNA Library Preparation: Invitrogen Corp
 CDNA Library Arrayed by: the I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: L1AM14280 row: h column: 12
 High quality sequence stop: 650.
 Location/Qualifiers
 1. 918
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6720636"
 /cell_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231,
 hTERT-HME1, LNCaP"
 /lab_host="EMDH10B"
 /clone_lib="MAPcL"
 /notes="Vector: pCMV-SPORT6; Site 1: EcoRV; Site 2: Not I;
 Subtracted with brain, liver, lung, kidney and muscle.
 Directionally cloned. Priming method: oligo-dT. Average
 insert size: 1800 bp. Library amplification: 26,000 fold.
 Kristi A. Eglund, James J. Vincent, Robert Strausberg,
 Bungkok Lee & Ira Pastan: Discovery of new breast
 cancer genes encoding membrane and secreted proteins.
 Manuscript submitted."

ORIGIN

Alignment Scores:
 Pred. No.: 3,01e-203 Length: 918
 Score: 238.00 Matches: 251
 Percent Similarity: 99.60% Conservative: 0
 Best Local Similarity: 99.60% Mismatches: 0
 Query Match: 28.71% Indels: 1
 DB: 14 Gaps: 0
 US-09-916-849A-1 (1-829) x CA454560 (1-918)
 Qy 180 LysTyrcGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMet 199
 Db 3 AAGTATGAGTCTTTGGCCACGCTGTGTACAGAAATGGTGCTCAGTGGAGGACCCATG 62
 Qy 200 AsnIleSerIleLeuValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThr 219
 Db 63 AACATCTCCATCATCGTGACCGACCAAGATGACCAAGCCCAAGTTTACCCAGGACAC 122
 Qy 220 PheArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAla 239
 Db 123 TTCGAGGGAGTGTCTTAGAGGGAGTCTTACAGGACTTCTGTGATGAGAGTGCAGCC 182
 Qy 240 ThrAspGluAspAlaIleTyThrTyAsnGlyValValAlaTySerIleHisSer 259
 Db 183 ACGGATGAGGATGATGCACTACCTACCAATGGGTGGTTCCTTACTCCATCCATAGC 242
 Qy 260 GlnGluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIle 279
 Db 243 CAAGAACAAGACCCACACAGCCTCATGTTCCACATTACCCGAGACAGGACCATC 302
 Qy 280 SerValIleSerSerGlyLeuAspArgGluLysValProGluTyThrLeuThrIleGln 299
 Db 303 AGCGTCACTCTCCAGTGGCTGGACCGGAAAGTCCCTGAGTACACACTGACCATCCAG 362
 Qy 300 AlaThrAspMetAspGlyAspGlySerThrThrAlaValAlaValValGluIleLeu 319
 Db 363 GCCACAGACATGATGGGACGGCTCCACCCACCGACGAGTGGCAGTAGTGGAGATCCTT 422
 Qy 320 AspAlaAsnAspAlaProMetPheAspProGlnLysTyrcGluAlaHisValProGlu 339
 Db 423 GATGCCAATGACAATGCTCCCATGTTTGACCCAGAGTACGAGGCCCATGTGCTGAG 482
 Qy 340 AsnAlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSer 359
 Db 483 AATGCACTGGGCCATGAGGTGACAGGCTGACGGTCACTGATCTGGACGCCCACTCA 542
 Qy 360 ProAlaThrArgAlaThrTyThrLeuIleMetGlyAspAspGlyAspHisPheThrIle 379
 Db 543 CCAGCGGGGTGCCACTTACCTTATATGGGGGTGACACGGGACCACTTTTACATC 602
 Qy 380 ThrThrHisProGluSerAsnGlnGlyIleLeuThrThrArg-LysGlyLeuAspPheGln 399

```

Db      603 ACCACCCACCTGAGACCAACGAGGCTCCTGACACCAACGAGGAGGTTTGATTTTGA 662
QY      399 uAlaLysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLys 419
Db      663 GSCCAAAACACGACACACCTGCTGACCTGAGTGAAGTACCAACGAGGCCCCCTTTGTGCTGAA 722
QY      419 sLeuProThrSerThrAlaThrIleValValHis 430
Db      723 GTCCCAACCTCCACAGCCACCACCATAGTGTGCCAC 756

RESULT 4
BM923882
LOCUS   BM923882
DEFINITION AGENCOURT_6709933 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5760016
5', mRNA sequence.
ACCESSION BM923882
VERSION   BM923882.1 GI:19374261
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
REFERENCE 1 (bases 1 to 1016)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/.
JOURNAL   National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT   Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12806 row: f column: 17
High quality sequence stop: 703.

FEATURES
Source
1..1016
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5760016"
/lab_host="DH10B"
/clone_lib="NIH_MGC_116"
/notes="Organ: Pooled colon, kidney, stomach; Vector:
pCMV-SPOK16; Site: 1: NotI; Site 2: EcoRV (destroyed); RNA
source anonymous pool of 3 colons, age 26 yo male, 49 yo
female, 71 yo male colon; 46 yo male kidney, and pool of 2
stomachs, 62 yo male and 70 yo female. Library is
oligo-dT primed and directionally cloned (EcoRV site is
destroyed upon cloning). Average insert size 1.4 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gruber (Invitrogen). Research Genetics tracking code
023. Note: this is a NIH_MGC Library."

ORIGIN
Alignment Scores:
Pred. No.: 1,68e-200 Length: 1016
Score: 235.00 Matches: 248
Percent Similarity: 99.60% Conservative: 0
Best Local Similarity: 99.60% Mismatches: 0
Query Match: 28.35% Indels: 1
DB: 12 Gaps: 0

US-09-916-849A-1 (1-829) x BM923882 (1-1016)

QY      1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuGluValCysTrpLeuGln 20
Db      45 ATGGGCTCCCTCGTGACCTCTCGGCTCTCTCCCTCTCCAGGTTTGTGCTGCAC 104

```

```

QY      21 CysAlaLysSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
Db      105 TGC CGCGCTCCGAGCGCTGCCGCGCTCTT CAGGAGGCTGAAGTGAACCTTTGGAGCG 164
QY      41 GlyGlyAlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
Db      165 GGAGCGCGGAGCAGGAGCCCGCCAGCGCTGGGAAAGTATTTCATGGCTGCCCTGGG 224
QY      61 GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr 80
Db      225 CAAGAGCCAGCTCTGTTTAGCACTGATAATGATGACTTCACTGTGCGGAATGGCAGACA 284
QY      81 ValGlnGluArgArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
Db      285 GTC CAGGAAGAAGTCACTGAAGAAAGAAATCCATTAAGATCTTCCCATCAAAACGT 344
QY      101 IleLeuArgArgHisLysArgAspTrpValValAlaProIleSerValProGluAsnGly 120
Db      345 ATCTTACGAAAGACACAAGAGAGATTGGTGGTGTCTCCATATCTGTCCCTGAAATGGC 404
QY      121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
Db      405 AAGGTCCTCCCTTCCCGCAGAGACTGAATCAGTCAAGTCTAATAAGATGAGACACCAAG 464
QY      141 IlePheTyrSerIleThrGlyProGlyValAlaAspSerProGluGlyValPheAlaVal 160
Db      465 ATTTCTACAGCATCACGGGCGCGGCGGAGACAGCCCCCTGAGGGTGTCTTCGCTGTA 524
QY      161 GluLysGluThrGlyTrpLeuLeuLeuAsnLysProLeuAspArgGluGluLysAlaLys 180
Db      525 GAGAAGGAGACAGGCTGTTGTTGTTGAATAAGCACTGGACCGGAGAGATTGCCAAG 584
QY      181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
Db      585 TATGAGCTCTTTGGCCACGCTGTGT CAGAAATGGTCCCTCAGTGGAGGACCCCATGAAC 644
QY      201 IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThr-Ph 220
Db      645 ATCTCCATCATCTGTGACCGACCAAGATGACCAAGCCCAAGTTTACCAGGACACCNNT 704
QY      220 eArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
Db      705 CCGAGGAGAGTGTCTTAGAGGAGTCTTACCAGGAGTACTTCTGTGATGAGGTGACGAC 764
QY      240 rAspGluAspAspAlaIleTyrThr 248
Db      765 AGATGAGGATGATGCCATCTACACC 789

RESULT 5
CA489015
LOCUS   CA489015
DEFINITION AGENCOURT_10809410 MAPcL Homo sapiens cDNA clone IMAGE:6721249 5',
mRNA sequence.
ACCESSION CA489015
VERSION   CA489015.1 GI:24951806
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
REFERENCE 1 (bases 1 to 875)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Kristi A. Eglund, Ira Pastan
cDNA Library Preparation: Invitrogen Corp
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

```

Plate: LLAM14282 row: b column: 01
High quality sequence stop: 647.

FEATURES

source

1. .875
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/db_xref="IMAGE:6721249"
/cell_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231,
hTERT-RME1, LNCaP"
/lab_host="EMD110B"
/clone_lib="MAPCL"
/notes="Vector: pCMV-SPORT6; Site 1: EcoRV; Site 2: Not I;
Subtracted with brain, liver, lung, kidney and muscle.
Directionally cloned. Priming method: oligo-dT. Average
insert size: 1800 bp. Library amplification: 26,000 fold.
Kristi A. Eglund, James J. Vincent, Robert Strausberg,
Bungkook Lee & Ira Pastan: Discovery of new breast
cancer genes encoding membrane and secreted proteins.
Manuscript submitted."

ORIGIN

Alignment Scores:
Pred. No.: 2,86e-194 Length: 875
Score: 228.00 Matches: 228
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 27.50% Indels: 0
DB: 14 Gaps: 0

US-09-916-849A-1 (1-829) x CA489015 (1-875)

QY 576 ThrSerProPheGlnAlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluVal 595
DB 35 ACCTCCCTTTCAGGCCAGCTCAGATGATCTAGATCTACTGGCGGAGGTC 94
QY 596 AsnGluGlyAspThrValValLeuSerLeuLysLeuLysPheLeuLysGlnAspThrTyr 615
DB 95 AACGAGAGAGTGACACAGTGTCTTGTCTCCTGAAGAGTCTCTGAAGCAGGATACATAT 154
QY 616 AspValHisLeuSerLeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAla 635
DB 155 GAGTGCACCTTCTCTGCTGACCTGCAACAGACAGCTGACGTGTATCAGGCC 214
QY 636 ThrValCysAspCysHisGlyHisValGluThrCysProGlyProThrPheGlyGlyPhe 655
DB 215 ACTGTGTGCGACTGCCATGCCATGTTCGAAACCTGCGCCCTGGACCTTGAAGGAGGTTTC 274
QY 656 IleLeuProValLeuGlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeu 675
DB 275 ATCCTCCCTGTGTGGGGGTGTCCTGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTTTG 334
QY 676 LeuValArgLysLysArgLysLysLysGluProLeuLeuLeuProGluAspAspThrArg 695
DB 335 TTGGTGAGAGAGAGCGGAGATCAGAGGCCCTCTCTACTCCAGAGATGACACCGT 394
QY 696 AspAsnValPheTyrTyrGlyGluGlyGlyGlyGlyGluAspGlnAspTyrAspIle 715
DB 395 GACAACGCTCTTCTACTATGCGAAGAGGGGGTGGCGAAGAGCAGCAGACTATGACATC 454
QY 716 ThrGlnLeuHisArgGlyLeuGluAlaArgProGluValValLeuArgAsnAspValAla 735
DB 455 ACCAGCTCCAGAGTCTGGAGGCGAGGCCGAGGTGTCTCTCGCATGACGTGGCA 514
QY 736 ProThrIleIleProThrProMetTyrArgProArgProAlaAsnProAspGluIleGly 755
DB 515 CCAACATCATCCGACACCATGTATCCGTCTCTCGGCCAGCCCAACCCAGATGAATCGGC 574
QY 756 AsnPheIleIleGluAsnLeuLysAlaAlaAsnThrAspProThrAlaProThrAsp 775
DB 575 AACTTTAATATGAGAACTGAAGCGGCTTAACACAGACCCCAACAGCCCGCCCTACGAC 634
QY 776 ThrLeuLeuValPheAspTyrGluGlySerGlySerAspAlaAlaSerLeuSerLeu 795

DB 635 ACCCTCTGTGTCTGACTATGAGGCGAGGGTCCGACCGCGTCCCTGAGCTCCCTC 694
QY 796 ThrSerSerAlaSerAspGlnAsp 803
DB 695 ACCTCTCTCGGCTCCGACCAAGAC 718
RESULT 6
BM545620
LOCUS
DEFINITION BM545620 1122 bp mRNA linear EST 20-FEB-2002
5', mRNA sequence.
ACCESSION BM545620
VERSION BM545620.1 GI:18777871
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (Bases 1 to 1122)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-rcmail.nih.gov
Tissue Procurement: Invitrogen
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12358 row: b column: 18
High quality sequence start: 108
High quality sequence stop: 742.
Location/Qualifiers
1. .1122
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5587889"
/lab_host="DH10B"
/clone_lib="NIH_MGC_125"
/notes="Organ: ovary (pool of 3); Vector: pCMV-SPORT6;
Site 1: EcoRV (destroyed); Site 2: NotI; RNA source pool
of three ovaries, from females ranging in age from 38 to
49 yo. Library is oligo-dT primed and directionally cloned
(EcoRV site is destroyed upon cloning). Average insert
size 2.1 kb, insert size range 1-3.5 kb. Library is
normalized and enriched for full-length clones and was
constructed by C. Gruber (Invitrogen). Research Genetics
tracking code 036."

ORIGIN

Alignment Scores:
Pred. No.: 3.7e-194 Length: 1122
Score: 228.00 Matches: 244
Percent Similarity: 99.59% Conservative: 0
Best Local Similarity: 99.59% Mismatches: 1
Query Match: 27.50% Indels: 1
DB: 12 Gaps: 0

US-09-916-849A-1 (1-829) x BM545620 (1-1122)

QY 585 AspAspSerAspIleTyrTrpThrAlaGluValAsnGluGlyAspThrValValLeu 604
DB 120 GATGACTCAGACATCTACTCGACGCGGAGGTCAACGAGGAGGTGACCA-GTGGTCTTG 178
QY 605 SerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSerLeuSerAspHis 624
DB 179 TCCCTGAGAGAGTTCCTGAAGCAGGATACATATGACGTGCACCTTCTCTGTCTGACCAT 238
QY 625 GlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCysHisGlyHisVal 644

Db	239	GGCAACAAGACAGCAGCTACGGTGATCAGGCCCACTGTGTGGCACTGCCATGCCCATGTC	298
QY	645	GluThrCysProGlyProTrpPlySerGlyGlyPheIleLeuProValLeuGlyAlaValLeu	664
Db	299	GAACCTGCCTCGGACCTTGAAGAGGAGGTTTCATCTCCTGTGCTGGGGCTGTCTCTG	358
QY	665	AlaLeuLeuPheLeuLeuValLeuLeuLeuValArgIlysIysArgIlysIleIys	684
Db	359	GCTCTGCTGTTCTCTCTGCTGGTCTGTTTTGTGTGAAGAAGACGGAAGATCAAG	418
QY	685	GluProLeuLeuProGluLeuProGluLeuProGluLeuProGluLeuProGluLeu	704
Db	419	GAGCCCTCTCTACTCCAGAGATGACACCCGTGCAACGTTCTTCTACTATGCGAAGAG	478
QY	705	GlyGlyGlyGluGluLeuGluLeuGluLeuGluLeuGluLeuGluLeuGluLeu	724
Db	479	GGGGGTGGCGAAGAGGACGAGACTATGACATCATCCAGCTCCACCGAGTCTGTGGAGGC	538
QY	725	ArgProGluValValLeuArgAsnAspValAlaProThrIleLeuProThrProMetTyr	744
Db	539	AGCGCGGAGTGGTCTCTCGCANTGACGTGCACCAACCATCATCCGACACCCATGTAC	598
QY	745	ArgProArgProAlaAsnProAspGluIleGlyAsnPheIleLeuLeuLeuLeuLeu	764
Db	599	CGTCTCTCGGCCACGCAACCCAGATCAATCGCACTTTTATATTGGAACCTTGAAGGCG	658
QY	765	AlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPheAspTyrGluGly	784
Db	659	GCTACACAGACCCACAGCCCGCCCTACGACACCTCTTTGGTTTCGATATGAGGCG	718
QY	785	SerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSerAspGlnAspGln	804
Db	719	AGCGGCTCGACGCGCGCTCTGAGTCTCCTACCTCCCTCCGCTCCGACCAAGACCAA	778
QY	805	AspTyrAspTyrLeuAsnGluTyrGlySerArgPheIysIysLeuAlaAspMetTyrGly	824
Db	779	GATTACGATTATCTCAACGAGTGGCGGACCGCTTCAAGAGCTGGCAGCATGTACGCT	838
QY	825	GlyGlyGluLeuAspAsp 829	
Db	839	GGCGGGGAGGACGAC 853	
RESULT 7			
BG403092			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			

CA455046 887 bp mRNA linear EST 12-NOV-2002
 AGENCOURT 10714428 MAPcL Homo sapiens cDNA clone IMAGE:6722552 5',
 mRNA sequence.
 ACCESSION CA455046
 VERSION CA455046.1 GI:24905371
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 887)
 NIH-MGC <http://mgi.nci.nih.gov/>.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgapbs@mail.nih.gov
 Tissue Procurement: Kristi A. Eglund, Ira Pastan
 CDNA Library Preparation: Invitrogen Corp
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM14285 row: h column: 08
 High quality sequence stop: 594.
 FEATURES
 Location/Qualifiers
 1..887
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6722552"
 /cell_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231,
 HERT-MB1, LNCaP"
 /lab_host="EMD108"
 /clone_lib="MAPcL"
 /notes="Vector: pCMV-SPORT6; Site 1: EcoRV; Site 2: Not I;
 Subtracted with brain, liver, lung, kidney and muscle.
 Directionally cloned. Priming method: oligo-dT. Average
 insert size: 1800 bp. Library amplification: 25,000 fold.
 Kristi A. Eglund, James J. Vincent, Robert Strausberg,
 Bungkook Lee & Ira Pastan. Discovery of new breast
 cancer genes encoding membrane and secreted proteins.
 Manuscript submitted."

ORIGIN
 Alignment Scores: 1.15e-190 Length: 887
 Pred. No.: 224.00 Matches: 251
 Score: 99.21% Conservative: 0
 Percent Similarity: 99.21% Mismatches: 1
 Best Local Similarity: 27.02% Indels: 2
 Query Match: 27.02% Gaps: 0
 DB: 14

US-09-916-849A-1 (1-829) x CA455046 (1-887)

QY 287 AspArgGluValProGluTyrThrLeuThrIleGlnAlaThrAspMetAspGlyAsp 306
 Db 3 GACCGGAAAAGTCCTCGTAGTACACTGACCTCAGCCAGGCCACACACATGGATGGGAC 62
 QY 307 GlySerThrThrAlaValAlaValGluLeuAlaAspAsnAlaPro 326
 Db 63 GGTTCACACACACCGGAGGAGTAGGAGATCCTTGATGCCAATGCAATGCTCCC 122
 QY 327 MetPheAspProGlnLysTyrGluAlaHisValProGluAsnAlaValGlyHisGluVal 346
 Db 123 ATGTTTACCCCCAGAGTAGCAGGCCATGTGCTGAGATGTCAGTGGCCATGAGGTG 182
 QY 347 GlnArgLeuThrValThrAspLeuAspAlaProAsnSerProAlaThrArgAlaThrTyr 366
 Db 183 CAGAGGCTGACGTCCTACTGATCTGAGCGCCCCCACTCAGCGCGTGGTCCACCTAC 242
 QY 367 LeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThrHisProGluSerAsn 386

Db 243 CTTATCATGGCGGTGACGACGGGGACCATTTTACCATCACACCACCTCGAGACAC 302
 QY 387 GlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAlaLysAsnGlnHisThrLeu 406
 Db 303 CAGGGCATCTGCACACACGAGAGGGTTTGATTGAGGCCAAAACACGACACCCCTG 362
 QY 407 TyrValGluValThrAsnGluAlaProPheValLeuLeuLysLeuProThrSerThrAlaThr 426
 Db 363 TACGTTGAAGTGCACACGAGGCCCTTTGTGCTGAAGCTCCCACTCCACGACACC 422
 QY 427 IleValValHisValGluAspValAsnGluAlaProValPheValProProSerLysVal 446
 Db 423 ATAGTGTTCACGTGGAGGATGTGAATGAGGCACCTGTGTTGTCCTCCAAAGTC 482
 QY 447 ValGluValGlnGluGlyIleProThrGlyGluProValCysValTyrThrAlaGluAsp 466
 Db 483 GTTGAGGTCCAGGAGGACATCCCACTCGGGGAGCCTGTGTGCTCTACACTGCAGAGAC 542
 QY 467 ProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArgAspProAlaGlyTyrLeu 486
 Db 543 CTTGACAGAGAGATCAAAAGATCAGCTACCGCATCTGAGACACCCAGCAGGTTGGCTA 602
 QY 487 AlaMetAspProAspSerGlyGlnValThrAlaValGlyThrLeuAspArgGluAspGlu 506
 Db 603 GCCATGGACACAGACAGTGGCAGGTGCACAGCTGTGGCACCCTCGACCTGAGGATGAG 662
 QY 507 GlnPheValArg-AsnAsnIleTyrGluValMetValLeuAlaMetAspAsnGlySerPr 526
 Db 663 CAGTTTGTGAGGAACAC-ATCTATGAAGTCATGGCTTGTGGCCATGACANTGAGAGCCC 721
 QY 526 cProThrThrGlyThrGlyThrLeuLeuThrLeu 538
 Db 722 TCCACCACTGGCAGCGGAACCTTCTGCTAACACTG 758

RESULT 9
 BU543746 865 bp mRNA linear EST 13-SEP-2002
 LOCUS AGENCOURT 10324520 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6575911
 DEFINITION 5', mRNA sequence.
 ACCESSION BU543746
 VERSION BU543746.1 GI:22854229
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 865)
 NIH-MGC <http://mgi.nci.nih.gov/>.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgapbs@mail.nih.gov
 Tissue Procurement: DCTD/DTF
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LICM2773 row: j column: 07
 High quality sequence stop: 654.
 Location/Qualifiers
 1..865
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6575911"
 /tissue_type="carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_40"
 /notes="Organ: prostate; Vector: pOTB7; Site 1: XhoI;
 Site 2: EcoRI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the

FEATURES
 source

following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:

Pred. No.: 5,63e-188 Length: 865
Score: 221.00 Matches: 221
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 26.66% Indels: 0
DB: 13 Gaps: 0

US-09-916-849A-1 (1-829) x BU543746 (1-865)

QY 178 IleAlaIysTyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAsp 197
Db 3 ATTGCCAAGTATGAGCTCTTTGGCCACGCTGTGTACAGAAATGGTCCATCGAGGAGAC 62
QY 198 ProMetAsnIleSerIleValThrAspGlnAsnAspHisLysProLysPheThrGln 217
Db 63 CCATCAACATCTCCATCATCTGTCGACCGACGAGATGACCAAGCCCAAGTTTACCCAG 122
QY 218 AspThrPheArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnVal 237
Db 123 GACACCTTCGAGGGAGTCTCTAGAGGAGTCTTACCAGAGTCTCTGTATGAGGATG 182
QY 238 ThrAlaThrAspGluAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIle 257
Db 183 ACAGCCACGGATGAGATGATGCCATCTACCTACATGAGGGTGTGCTTACTCCATC 242
QY 258 HisSerGlnGluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGly 277
Db 243 CATAGCAAGAACCAAGGACCCACACGACCTCATGTTCCACCATTCACCGAGCACAGGC 302
QY 278 ThrIleSerValIleSerGlyLeuAspArgGlyValProGluTyrThrLeuThr 297
Db 303 ACATCAGCGTCATCTCCAGTGCCTCGACCGGAAAGTCCCTGAGTACATCTGACC 362
QY 298 IleGlnAlaThrAspMetAspGlyAspGlySerThrThrAlaValAlaValGlu 317
Db 363 ATCCAGGCCACAGACATGATGGGACGCTCCACCCACCGCAGTGGCAGTAGTGAG 422
QY 318 IleLeuAspAlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisVal 337
Db 423 ATCCCTTGATGCCAATGACATGCTCCCATGTTGACCCGCCAGAGTACGAGGCCCATGTG 482
QY 338 ProGluAsnAlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaPro 357
Db 483 CCTGAGATGCAATGGCCATGAGTGCAGAGGCTGACGCTCACTGATCTGACGCCCCC 542
QY 358 AsnSerProAlaTyrArgAlaThrTyrLeuIleMetGlyGlyAspAspHisPhe 377
Db 543 AACTCACCAGCGTGGCGTCCACCTACCTTATCATGGCGGTGACGAGCGGACCATTTT 602
QY 378 ThrIleThrThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAsp 397
Db 603 ACATCACCACCCACCTGAGAGCAACACGAGGATCTCTGACCAACACGAGGAGGTGTGAT 662
QY 398 Phe 398
Db 663 TTT 665

RESULT 10

BM715802

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

BM715802 703 bp mRNA linear EST 28-FEB-2002
UI-E-EJO-ahj-m-06-0-UI.r2 UI-E-EJO Homo sapiens cDNA clone
UI-E-EJO-ahj-m-06-0-UI 5', mRNA sequence.
BM715802
BM715802.1 GI:19029060
EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (Bases 1 to 703)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery

JOURNAL

Genome Res. 6 (9), 791-806 (1996)

MEDLINE

97044477

PUBMED

889548

COMMENT

Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9585
Email: bento-soares@uiowa.edu

Tissue Procurement: Dr. Gregg Hageman
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seq primer: M13 Reverse.

FEATURES

source

Location/Qualifiers
1..703
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-E-EJO-ahj-m-06-0-UI"
/tissue_type="fetal eyes, lens, eye anterior segment,
optic nerve, retina, Retina Foveal and Macular, RPE and
Choroid"
/dev_stage="fetal and adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-E-EJO"
/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site: 1: EcoR I; Site 2: Not I;
UI-E-EJO is a subtracted cDNA library constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pT73-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for this library are: fetal eyes,
AGATCAAGA: lens, CGATTAGCGA: eye anterior segment,
AATCCCGCAT: optic nerve, CCATTAGCTG: retina, CCGCG,
Foveal and Macular, GTCC; RPE and Choroid, ACCTA. This
library was created for the program, Gene Discovery in the
Visual System, supported by National Eye Institute (NEI)."

ORIGIN

Alignment Scores:

Pred. No.: 9,05e-182 Length: 703
Score: 214.00 Matches: 227
Percent Similarity: 99.56% Conservative: 0
Best Local Similarity: 99.56% Mismatches: 0
Query Match: 25.81% Indels: 1
DB: 12 Gaps: 0

US-09-916-849A-1 (1-829) x BM715802 (1-703)

QY

336 HisValProGluAsnAlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAsp 355

Db

6 CATGTGCTGAGATGTCAGTGGGCGCATAGGTGACAGAGGCTGACGCTCACTGATCTGGAC 65

QY

356 AlaProAsnSerProAlaTyrArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAsp 375

Db

66 GCCCCCAACTCACAGCGTGGCGTGCCACCTTATCATGTGGCGGTGACGAGGGGAC 125

376 HisPheThrIleThrThrHisPro-GluSerAsnGlnGlyIleLeuThrThrArgLysGI 395
126 CATTTTACCATCACCACTTGAGACCAACAGGCGATCTGTGACCAACAGAGGG 185
395 YLeuAspPheGluAlaLysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaPr 415
186 TTGGATTGAGCCCAAAACAGACACCTGTACGTGAGTGAGTGCACCAACAGGCCCC 245
415 oPheValLeuLeuLeuProThrSerThrAlaThrIleValValHisValGluAspValAs 435
246 TTTTGTGCTGAAGCTCCCAACCTCCACAGCCACCATAGTGTGCTCCAGTGGAGATGTGAA 305
435 nGluAlaProValPheValProSerLysValValGluValGluValGlyIleProTh 455
306 TGAGGCACCTGTGTTGTCCTCCACCTCCAAAGTCGTTGAGGTCAGGAGGGCATCCCCAC 365
455 rGlyGluProValCysValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSe 475
366 TGGGGAGCCTGTGTGTCTACACTGCAGAGACCCCTGACAAAGGAGATCAAAAGATCAG 425
475 rTyrArgIleLeuArgAspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGluVa 495
426 CTACCCGATCTCTGAGAGACCCAGCAGCGTGGCTAGCCATGGACCCAGAGAGTGGCAGGT 485
495 lThrAlaValGlyThrLeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGI 515
486 CACAGCTGTGGGCACCTCGACCGTGGAGTGGAGTGTGAGGAGAACACATCATGA 545
515 uValMetValLeuAlaMetAspAsnGlySerProThrThrGlyThrGlyThrLeuLeu 535
546 AGTCATGCTCTTGCCATGTGACCAATGAAGCCCTCCACCACTGGCAGCGGAACCTTCT 605
535 uLeuThrLeuLeuAspValAsnAspHisGlyProValProGluProArgGlnIleThrII 555
606 GCTAACACTGATTGATGTCATGACCATGCCCAGTCCCTGAGCCCGTCAGATCACCAT 565
555 eCysAsnGlnSerProValArg 562
666 CTGCAACCAAGCCCTGTGGGT 687

RESULT 11
BUI68948 1006 bp mRNA linear EST 04-SEP-2002
LOCUS AGENCOURT_8110724 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6266745
DEFINITION 5', mRNA sequence.
ACCESSION BUI68948.1 GI:22682932
VERSION BUI68948
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1006)
NIH-MGC <http://mgi.nci.nih.gov/>
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTF
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LILNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LILNL at:
<http://image.llnl.gov>
Plate: LLCM2435 row: p column: 10
High quality sequence stop: 607.
Location/Qualifiers
1..1006
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"

FEATURES
source
CB240552/c
LOCUS
DEFINITION
UI-CF-PNO-afv-b-02-0-UI.s1 UI-CF-PNO Homo sapiens cDNA clone
UI-CF-PNO-afv-b-02-0-UI.3', mRNA sequence.

/clone="IMAGE:6266745"
/tissue type="melanotic melanoma, cell line"
/lab host="DH10B (phage-resistant)"
/clone lib="NIH_MGC_112"
/notes="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGACAGCAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
ScripScript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

ORIGIN
Alignment Scores: 1,31e-181 Length: 1006
Pred. No.: 214.00 Matches: 214
Score: 214.00
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 25.81% Indels: 0
DB: 13 Gaps: 0

US-09-916-849A-1 (1-829) x BUI68948 (1-1006)

QY 320 AsnAlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSer 359
DB 2 GATGCCAATGCAATGCTCCCATGTTGACCCCAAGTACGAGGCCCATGTGCTTGTGAG 61
QY 340 AsnAlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSer 359
DB 62 AATGCATGTGGCCATGAGTGCAGAGGCTGACGCTCATCTGTGACGCCCCCACTCA 121
QY 360 ProAlaTrpArgAlaThrTyrLeuIleMetGlyValAspAspGlyAspHisPheThrIle 379
DB 122 CCAGCGTGGGTGGCCACTTATCATGGCGGTGACACGGGACCACTTTTACATC 181
QY 380 ThrThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGlu 399
DB 182 ACCACCCACCTTGAGACCAACAGGCGCATCTGCAACACAGGAGGGTTTGGATTTGAG 241
QY 400 AlalysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLys 419
DB 242 GCCAAAAACAGCACACCTGTGACGTTGAAGTACCAACAGGAGGCCCTTTTGTGCTGAAG 301
QY 420 LeuProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProVal 439
DB 302 CTCCCAACCTCCACAGCACCATAGTGTGTCACGTGGAGATGTGATGAGGACCTGTG 361
QY 440 PheValProSerLysValValGluValGlnGlyIleProThrGlyGluProVal 459
DB 362 TTTGTCCACCTCCAAAGTCGTTGAGTCCAGAGGCGCATCCCACTGGGAGGCTGTG 421
QY 460 CysValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeu 479
DB 422 TGTGTCTACACTGCAGAGACCCCTGACAGGAGATCAAAAGATCAGCTACCGCATCTG 481
QY 480 ArgAspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGly 499
DB 482 AGAGACCCAGCAGGGTGGCTAGCCATGACCCACAGTGGGACAGTCCACAGCTGTGGGC 541
QY 500 ThrLeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeu 519
DB 542 ACCCTCGACCGTGGAGTGCAGCTTTGTGAGGAGCAACATCTATGAGTGAAGTCATGCTTG 601
QY 520 AlaMetAspAsnGlySerProThrThrGlyThrGlyThr 533
DB 602 GCCATGGACAATGGAAGCCCTCCCACTGGCAGCGGAACC 643

RESULT 12
CB240552/c
LOCUS
DEFINITION
UI-CF-PNO-afv-b-02-0-UI.s1 UI-CF-PNO Homo sapiens cDNA clone
UI-CF-PNO-afv-b-02-0-UI.3', mRNA sequence.

```

ACCESSION      CB240552
VERSION        CB240552.1
KEYWORDS       GI:28362196
SOURCE         EST.
ORGANISM       Homo sapiens (human)

REFERENCE
AUTHORS        Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE          Eukaryota; Metazoa; Primates; Catarrhini; Hominidae; Homo.
JOURNAL        1 (bases 1 to 745)
MEDLINE        Bonaldo,M.F., Lennon,G. and Soares,M.B.
PUBMED         Normalization and subtraction: two approaches to facilitate gene
                discovery
                Genome Res. 6 (9), 791-806 (1996)
                97044477
                8889548
COMMENT        Contact: McCray, PB
                McCray Lab
                University of Iowa
                2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
                Tel: 319 356 4866
                Fax: 319 356 7171
                Email: paul-mccray@uiowa.edu
                Tissue Procurement: Dr. M. J. Welsh, University of Iowa
                cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
                cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
                DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
                Clone Distribution: Researchers may obtain clones from Research
                Genetics (www.resgen.com) or from Open Biosystems
                (www.openbiosystems.com).
                Seq primer: M13 FORWARD
                POLYA=yes.

FEATURES
    source
        1..745
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="UI-CF-FNO-afv-b-02-0-UI"
            /tissue_type="Human Lung Epithelial cells"
            /lab_host="PH10B (Life Technologies) (T1 phage resistant)"
            /clone_lib="UI-CF-FNO"
            /note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
            modified polylinker; Site 1: EcoR I; Site 2: Not I;
            UI-CF-FNO is a subtracted cDNA library derived from two
            normalized Human lung epithelial cell libraries (EN1 and
            DU1) The library was subtracted according to according to
            Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
            1996. For additional information, contact:
            bentso-soares@uiowa.edu
            TAG_TISSUE=Human Lung Epithelial Cell Lines untreated LPS
            6hr to LPS 24h
            TAG_LIB=UI-CF-FNO
            TAG_SEQ=CTGCTCAGGT"

ORIGIN
Alignment Scores:
Pred. No.:      7,638-181      Length:      745
Score:          213.00      Matches:      213
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    25.69%      Indels:      0
DB:             14      Gaps:          0

US-09-916-849A-1 (1-829) x CB240552 (1-745)

QY      53  LysValPheMetGlyCysProGlyGlnGluProAlaLeuPheSerThrAspAsnAspAsp 72
DB      659 AAAGTATTTCATGGCTGCCTGGCCAGAGCCAGCTCTGTTAGCACTGATAATGATGAC 600
QY      73  PheThrValArgAsnGlyGluThrValGlnGluArgGlySerLeuLysGluArgAsnPro 92
DB      599 TTCACGTGCGGAATGGCAGACAGTCACGAGGAAGAGTCACTGAGGAAGAAGGAATCCA 540
QY      93  LeuLysIlePheProSerLysArgIleLeuArgArgHisLysArgAspTrpValValAla 112

```

```

Db      539 TTGAAGATCTTCCCATCCAAACGATATCTTACGAAGACACAAAGAGAGATGGGTGGTGGCT 480
QY      113 ProIleSerValProGluAsnGlyLysGlyProPheProGlnArgLeuAsnGlnLeuLys 132
Db      479 CCAATATCTGTCCTGAAATGGCAAGGTCCTTCCCCAGAGACTGAATCAGCTCAAG 420
QY      133 SerAsnLysAspArgAspThrLysIlePhePheThrSerIleThrGlyProGlyAlaAspSer 152
Db      419 TCTAATAAAGATAGAGACACCAAGATTTCTTACAGCATCAGGGGCGGGGGGAGAGACG 360
QY      153 ProProGluGlyValPheAlaValGluLysGluThrGlyTrpLeuLeuLeuAsnLysPro 172
Db      359 CCCCTCGAGGGTGCTTCGCTGTAGAGAAGGAGACAGCTGGTGTGGTGAATAAGCCA 300
QY      173 LeuAspArgGluGluIleAlaLysTyGluLeuPheGlyHisAlaValSerGluAsnGly 192
Db      299 CTGACCGGGAGAGATTGCCAAGTATGAGCTCTTTGGCCACGCTGTGTGAGAGAATGGT 240
QY      193 AlaSerValGluAspProMetAsnIleSerIleValThrAspGlnAsnAspHisLys 212
Db      239 GCCTCAGTGGAGAGACCCCATGAACATCTCCATCATCTGACCGACCAAGATGACCACAAG 180
QY      213 ProLysPheThrGlnAspThrPheArgGlySerValLeuGluGlyValLeuProGlyThr 232
Db      179 CCCAAGTTTACCCAGGACACCTTCGAGGGAGTGCTCTTAGAGGGAGTCTCTACCAGGTACT 120
QY      233 SerValMetGlnValThrAlaThrAspGluAspAlaIleTyrThrTyrAsnGlyVal 252
Db      119 TCTGTGATGCAGGTGCAGCCAGCATGAGATGCTCATCTACCTACCTACATGGGGTG 60
QY      253 ValAlaTyrSerIleHisSerGlnGluProLysAspPro 265
Db      59 GTTGCTTACTCCATCCATAGCAAGAACCAAGAGGACCCA 21

CB992509      880 bp      mRNA      linear      EST 01-MAY-2003
AGENCY        IMAGE:30330659 5', mRNA sequence.
ACCESSION     CB992509
VERSION       CB992509.1
KEYWORDS      EST.
SOURCE        Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 880)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Dr. Stefan Hansson
cDNA Library Preparation: Michael J. Brownstein (NHGRI) with help
and advice from Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDAM344 row: n column: 12
High quality sequence stop: 641.
Location/Qualifiers
1..880
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30330659"
/tissue_type="pre-eclamptic placenta"
/lab_host="DH10B Tona"
/clone_lib="NIH_MGC_148"
/note="Organ: placenta; Vector: pBluescriptR; Site 1:
all-XhoI; Site 2: BamH; Library is oligo-dT primed and

```



```

Db      404  AAGGGTCCCTTCCCCAGAGACGTAATCAGTCAAGTCTAATAAAGATAGACACCAAG 463
QY      141  IlePheTyrSerIleThrGlyProGlyAlaAspSerProGluGlyValPheAlaVal 160
Db      464  ATTTTCTACAGCATCAGCGGGCGGGGCGAGACAGCCCCCTTGGGTGCTTCGCTGTA 523
QY      161  GluGlyGluThrGlyTTrpLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 180
Db      524  GAGAAGGAGACAGCGTGGTGTGTTGTTGAATAAGCCACTGGACCGGGAGAGATTGCCAAG 583
QY      181  TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAen 200
Db      584  TATGACTCTTTGGCCACCGCTGTGTAGAGATGGTGCCTCAGTGAGACCCCATGAC 643
QY      201  IleSerIleIleValThrAspGlnAenAspHis 211
Db      644  ATCTCCATCATCGTGACCGACCAAGATGACCAC 676

RESULT 15
BM804940
LOCUS   BM804940
DEFINITION AGENCOURT_6495104 NIH_MGC_125 Homo sapiens cDNA clone IMAGE:5725677
5', mRNA sequence.
ACCESSION BM804940
VERSION   1
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 999)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgabbs@mail.nih.gov
          Tissue Procurement: Invitrogen
          cDNA Library Preparation: Life Technologies, Inc.
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLAM12716 row: 0 column: 22
          High quality sequence start: 18
          High quality sequence stop: 680.

FEATURES
source    Location/Qualifiers
          1..999
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:5725677"
            /lab_host="DH10B"
            /clone_lib="NIH_MGC_125"
            /notes="Organ: ovary (pool of 3); Vector: pCMV-SPORT6;
            Site: 1: EcoRV (destroyed); Site 2: NotI; RNA source pool
            of three ovaries, from females ranging in age from 38 to
            49 yo. Library is oligo-dT primed and directionally cloned
            (EcoRV site is destroyed upon cloning). Average insert
            size 2.1 kb, insert size range 1-3.5 kb. Library is
            normalized and enriched for full-length clones and was
            constructed by C. Gruber (Invitrogen). Research Genetics
            tracking code 036."

```

ORIGIN

```

Alignment Scores:
Pred. No.:      3,27e-176      Length:      999
Score:          208.00         Matches:    208
Percent Similarity: 100.00%    Conservative: 0
Best Local Similarity: 100.00% Mismatches:    0
Query Match:    25.09%        Indels:     0
Db:              12          Gaps:      0

```

```

US-09-916-849A-1 (1-829) x BM804940 (1-999)
QY      54  ValPheMetGlyCysProGlyGlnGluProAlaLeuPheSerThrAspAsnAspPhe 73
Db      118  GTATTTCATGGGCTGCCCCCTGGGCAAGAGCCAGCTCTGTTTAGCACTGATAATGATGACTTC 177
QY      74  ThrValArgAsnGlyGluThrValGlnGluArgSerLeuLeuLeuLeuLeuLeuLeuLeu 93
Db      178  ACTGTGGGAATGGCGAGACAGTCCAGGAAGAAGAGGTCACCTGAAGGAAAGAAATCCATTG 237
QY      94  LysIlePheProSerLysArgIleLeuArgArgHisLysArgAspTrpValValAlaPro 113
Db      238  AAGATCTTCCCATCCAAACGATATCTTACGAGACACACAGAGAGATTGGTGGTGGTCTCCA 297
QY      114  IleSerValProGluLeuAsnGlyLysGlyProPheProGlnArgLeuLeuLeuLeuLeu 133
Db      298  ATATCTGTCCCTGAAAATGGCAAGGGTCCCTTCCCCAGAGACTGAATCAGCTCAAGTCT 357
QY      134  AsnLysAspArgAspThrLysIlePheTyrSerIleThrGlyProGlyAlaAspSerPro 153
Db      358  AATAAGATAGACACCAAGATTTTCTACAGCATCAGGGGCCCGGGGCGAGACAGACGCCC 417
QY      154  ProGluGlyValPheAlaValGluLysGluThrGlyTrpLeuLeuLeuLeuLeuLeuLeu 173
Db      418  CCTGAGGGTGTCTTCGCTGTAGAGAGGAGACAGGCTGGTGTGTTGTTGAATAAGCCACTG 477
QY      174  AspArgGluGluIleAlaLysTyrGluLeuPheGlyHisAlaValSerGluAsnGlyAla 193
Db      478  GACCGGGAGGAGATGGCCAGTATGAGCTCTTTGGCCACGCTGTGTGAGAGAATGGTGCC 537
QY      194  SerValGluAspProMetAsnIleSerIleValThrAspGlnAsnAspHisLysPro 213
Db      538  TCAGTGGAGGACCCCATGAACATCTCCATCATCTGACCGACCAAGATGACCACCAAGCCC 597
QY      214  LysPheThrGlnAspThrPheArgGlySerValLeuGluGlyValLeuProGlyThrSer 233
Db      598  AAGTTTACCAGACACCTTCCGAGGGAGTGTCTTAGAGGGGAGTCTTACCAAGTACTTCT 657
QY      234  ValMetGlnValThrAlaThrAspGluAspAlaIleTyrThrTyrAsnGlyValVal 253
Db      658  GTGATGCAAGTGACAGCACCGGATGAGGATGATGCCATCTACACCTACCAATGGGGTGTT 717
QY      254  AlaTyrSerIleHisSerGlnGlu 261
Db      718  GCTTACTCATCCATAGCCAAAGAA 741

```

Search completed: September 23, 2004, 07:38:32
Job time : 5105 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 23, 2004, 01:28:20 / Search time 8137 Seconds
(without alignments)

4415.801 Million cell updates/sec

Title: US-09-916-849A-1

Perfect score: 829

Sequence: 1 MGLPRGPIASLLLLQVCWLQ.....NEWGRFKKLADMYGGGEDD 829

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Word size: 1

Total number of hits satisfying chosen parameters: 6934743

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-O=/cpn2.1/USPTO.spool.h/US09916849/runat.21092004.145835.7317/app_query.fasta.1.967
-DB=GenEmbl -QFMT=fastap -SUFFIX=oligo.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=quality -THRESHOLD=15 -MODS=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09916849@CGN.1.1.5239 @runat.21092004.145835.7317 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -FGAPOP=60
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pin.*
35: em.htg.rod.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	829	100.0	3171	6	AR270785 Sequence
2	829	100.0	3171	6	AX411115 Sequence
3	829	100.0	3171	6	AX823232 Sequence
4	829	100.0	3171	9	X63629 H.sapiens m
5	715	86.2	3219	6	AX382118 Sequence
6	627	75.6	3686	9	BC041846 Homo sapi
7	102	12.3	306	6	AX210689 Sequence
8	80	9.7	184945	9	AC099314 Homo sapi
9	80	9.7	186839	30	AL139227 Human chr
10	80	9.7	186911	9	AC126773 Homo sapi
11	72	8.7	54207	2	AF002024 Homo sapi
12	61	7.4	2106	4	BTCDHP
13	54	6.5	255	6	AX210739 Sequence
14	51	6.2	2993	10	BC052189
15	51	6.2	3187	10	MCADHP
16	37	4.5	316	10	AF177683 Rattus no
17	28	3.4	179743	2	AC144675 Rattus no
18	28	3.4	190915	2	AC132132 Mus muscu
19	28	3.4	219657	2	AC129611 Rattus no
20	28	3.4	281505	2	AC120695 Rattus no
21	27	3.3	435	4	AF033826 Sus scrof
22	27	3.8	179743	2	AC144675 Rattus no
23	23	2.8	241282	2	AC094652 Rattus no
24	21	2.5	522	6	AX393688 Sequence
25	17	2.1	1055	5	CHKCAMA16
26	17	2.1	2436	5	GGCADHB
27	17	2.1	3545	5	CHKLCAMR
28	17	2.1	4434	5	CHKCAM
29	16	1.9	1684	5	XLBCAD
30	16	1.9	2310	6	AX598677 Sequence
31	16	1.9	2699	5	XLEPCAD
32	16	1.9	3450	5	XLU04707
33	16	1.9	3750	5	XLXBCADH
34	16	1.9	5310	6	AX767362 Sequence
35	15	1.8	164	10	MMUVEX15
36	15	1.8	168	5	CHKCAMA15
37	15	1.8	204	10	RNECADHER
38	15	1.8	313	10	AF177680 Rattus no
39	15	1.8	839	11	BV049696
40	15	1.8	923	9	HSPCADHGN
41	15	1.8	1902	10	MMUVEX16
42	15	1.8	2173	5	XLMECAD
43	15	1.8	2269	9	HSECAD16
44	15	1.8	2269	9	HUMCDH1014
45	15	1.8	2486	10	MMUVOM

ALIGNMENTS

RESULT 1


```

1794 GCCCAGGTCCAGATGACTCAGACATCTACTGGACGGCAGAGTCAACAGGAAGGTGAC 1853
QY ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
1854 ACATGGTCTTCTCCCTGAGAGTCTCTGAGCAGATACATATGAGTGCACCTTCT 1913
QY LeuSerAspHisGlyAsnLysGluLeuThrValLeuLysLysLys 640
1914 CTGTCTGACCATGCAACAAGAGCAGCTGACGGTCTGCTGCTGCTGCTGCTGCTGCT 1973
QY HisGlyHisValGluThrCysProGlyProLysGlyGlyPheLeuLeuProValLeu 660
1974 CATGCCATGTCGAAACCTGCCCTGACCTCGAAGAGGTTCATCTCTCCCTGCTG 2033
QY GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysLys 680
2034 GGGCTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2093
QY ArgLysLysLysGluProLeuLeuLeuProGluAspThrArgAspAsnValPheTyr 700
2094 CGGAAGATCAAGAGCCCTCTCTACTCCAGAGATGACACCCGCTGACACGCTTCTAC 2153
QY TyrGlyGluGluGlyGlyGlyGluGluAspGlnAspTyrAspPheLeuHisArg 720
2154 TATGCCAAGAGGGGGTGGCGAAGAGGACAGGACTATGACATCACCCAGCTCCACCA 2213
QY GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleLeuPro 740
2214 GGTCTGAGGCCAGCCGAGAGTGGTCTCCGAAAGAGTGGAGTGGAGTGGAGTGGAG 2273
QY ThrProMetTyrArgProArgProAlaAsnProAspGluLeuGlyAsnPheIleLeuGlu 760
2274 ACACCATGTACCGTCTAGCCAGCCACCCAGATGAAATCGCACTTTATATTTAG 2333
QY AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
2334 AACCTGAGCGGGCTAACAGACCCACCCAGCCGCCCTTACGACACCTCTTGTGTTC 2393
QY AspTyrGluGlySerGlySerAspAlaAsnLeuSerLeuSerLeuSerLeuSerLeuSer 800
2394 GACTATGAGCGAGCGGCTCCGACGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2453
QY AspGlnAspGlnAspTyrAspTyrLeuAsnGluTyrGlySerArgPheLysLysLeuAla 820
2454 GACCAAGACCAAGATTACGATTATCTGACAGTGGGGGAGCGGCTTCAAGAAGCTGGCA 2513
QY AspMetTyrGlyGlyGlyGluAspAsp 829
2514 GACATGTACGGTGGCGGGAGGAGCAG 2540

```

```

RESULT 2
AX411115 3171 bp DNA linear PAT 14-JUN-2002
LOCUS Sequence 3762 from Patent WO0229103.
DEFINITION AX411115
ACCESSION AX411115
VERSION AX411115.1 GI:21443820
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J. G.
TITLE Gene expression profiles in liver cancer
JOURNAL Patent: WO 0229103-A 3762 11-APR-2002;
GENE LOGIC INC (US)
FEATURES
source
1. 3171
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="EMBL/GenBank Accession No. X63629"

```

```

ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 3171
Score: 829.00 Matches: 829
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-09-916-849A-1 (1-829) x AX411115 (1-3171)
QY 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuGlnValCysThrLeuGln 20
DB 54 ATGGGGCTCCCTCGTGACCTCTCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 113
QY 21 CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
DB 114 TGGCGGCTCCCGAGCGCTGCGGGCGGCTCTTCAGGAGGCTGAATGAGCTTGGAGCG 173
QY 41 GlyGlyAlaGluGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
DB 174 GGAGGCGCGAGGAGGAGCGCGCGCGCTGGGAGAGTATTCATGGGCTGCCCTGGG 233
QY 61 GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr 80
DB 234 CAAGAGCCACTCTGTTAGCACTGATATGATGACTTCTCTGCGGATGGCGAGACA 293
QY 81 ValGlnGluArgSerLeuLysGluArgAsnProLeuLysLysPheProSerLysArg 100
DB 294 GTCCAGGAAAGAGGTCACTGAGGAAAGAAATCCATTGAAGATCTTCCCATCCAAAGT 353
QY 101 IleLeuArgArgHisLysArgAspTyrValValAlaProIleSerValProGluAsnGly 120
DB 354 ATCTTACGAGACACAAGAGAGATTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 413
QY 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
DB 414 AAGGGTCCCTTCCCGAGAGACTGAATCAGCTCAAGTCTTAATAAGATAGAGACCAAG 473
QY 141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProProGluGlyValPheAlaVal 160
DB 474 ATTTCTACAGCATACGCGGGCGGGGCGAGACAGCCCGCTTACCGGCTGCTCTGCTGTA 533
QY 161 GluLysGluThrGlyTrpLeuLeuLeuAsnLysProLeuAspArgGluLeuLysLys 180
DB 534 GAGAAGGAGACAGCGCTGGTGTGTTGTAATAAGCCACTGGACCGGGAGGAGATTGCCAAG 593
QY 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
DB 594 TATGAGCTCTTTGGCCACGCTGTCTCAGAGATGGTGCCTCAGTGGAGGACCCCATGAC 653
QY 201 IleSerIleLeuValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
DB 654 ATCTCCATCATCTGACCGAGAGTACCAAGCCCAAGTGTGCTTACTCCATCCAGACCTTC 713
QY 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
DB 714 CGAGGAGTGTCTTAGAGGAGTCTCCTACAGGATCTTCTGTGATGAGGAGGAGGAGGAG 773
QY 241 AspGluAspAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln 260
DB 774 GATGAGGATGATGCCATCTACACCTACATGGGGTGGTGGTGGTGGTGGTGGTGGTGG 833
QY 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrLysSer 280
DB 834 GAACCAAGGAGCCCAACGACCTCATGTTCACTTCCGAGGAGGAGGAGGAGGAGGAGGAG 893
QY 281 ValLysSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
DB 894 GTCATCTCCAGTGGCTGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 953
QY 301 ThrAspMetAspGlyAspGlySerThrThrAlaValAlaValAlaValGluLeuAsp 320

```

954	Db	ACAGACATGGATGGGACGGCTCCACCACCAGCGAGTGGCAGTAGTGGAGATCCTTGGAT	1013
321	Qy	AlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn	340
1014	Db	GCCAAATGACAAATGCTCCCATGTTTGGACCCCCAGAAAGTACGAGGCCCATGTGCCTGAGAAT	1073
341	Qy	AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro	360
1074	Db	GCATGGGCCCATGAGGTGCAGAGCTGACGGTCACTGATCTGGACGCCCACTACCA	1133
361	Qy	AlaTrpArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr	380
1134	Db	CGGTGGGTGCCACTTACCTTATCATGGGCGGTACGACGGGACCAATTTACATCACC	1193
381	Qy	ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla	400
1194	Db	ACCACACCTTGAGAGCAACCGAGGCATCTGTACACACAGGAAGGTTTGGATTTAGGCC	1253
401	Qy	LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu	420
1254	Db	AAAAACACGACACACCCCTGTAGTTGAAGTACCAACGAGGCCCTTTGTGCTGAAGCTC	1313
421	Qy	ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe	440
1314	Db	CCAACTCCACAGCCACCATAGTGGTCCACGTGGAGGATGTGAATGAGGCACCTGTGTTT	1373
441	Qy	ValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys	460
1374	Db	GTCCCCACCTCCAAAGTCGTGAGGTCCAGAGGGCATGCCCACTGGGAGCCTGTGTGT	1433
461	Qy	ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuAsq	480
1434	Db	GTCTACTCCAGAGACCCCTGCACAGGAGAATCAAAAGATCAGCTACCGCATCCTGAGA	1493
481	Qy	AspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr	500
1494	Db	GACCCAGCAGGGTGGCTAGCCATCGAACCCACAGAGTGGCAGGTTCACAGTGTGGGCACC	1553
501	Qy	LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla	520
1554	Db	CTCGACCGTGAGATGAGCAGTTTGTGAGAAACAATCTATGAAGTCATGGTCTGGCC	1613
521	Qy	MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAsp	540
1614	Db	ATGGCAATGGAAAGCCCTCCACCACTGGCAGCGGAACCCCTTCGTGTAACACTGATTGAT	1673
541	Qy	ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro	560
1674	Db	GTCAACGACATGGCCCGATCCCTGTAGCCCGGTACATCACCATCTGCAACCAAGACCTT	1733
561	Qy	ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln	580
1734	Db	GTGGCCACATGCTGAACATCACGGAAGGACCTGTCTCCACACACCTCCCTTTCAG	1793
581	Qy	AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGlyAsp	600
1794	Db	GCCACGCTCACAGATGACTCAGACATCTACTGGACGCGCAGAGGTCAACAGAGAAGTGAC	1853
601	Qy	ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer	620
1854	Db	ACATGGTCTGTTCCTGAAGAAGTTCCTGAAGCAGGATACATATGACGTGCACCTTCT	1913
621	Qy	LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys	640
1914	Db	CTGTCTGACATGGCAACAAAGAGCAGTGCACGGTATCAGGGGCCCATGTGTGCGACTGC	1973
641	Qy	HisGlyHisValGluThrCysProGlyProTrpLysGlyGlyPheIleLeuProValLeu	660
1974	Db	CATGGCCATGTGAACCTGCCCTGGACCTGGAAAGGAGGTTTATCTCTCCTGTGTGTG	2033
661	Qy	GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysLys	680

D	b	2034	GGGCGTGTCTCTGGCTCTGCTCTTTCTCTGCTGTGTCTGCTTTTTTGTTGGTGAGAAAGAAG	2093
Q	y	681	ArgLysIleLysGluProLeuLeuLeuProGluAspThrArgAspAenValPheTyr	700
D	b	2094	CGRAGATCAAGAGGCCCTCTACTCCAGAAAGATCACCCCGTGACAACGTCCTTCTAC	2153
Q	y	701	TyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArg	720
D	b	2154	TATGGCGAAGAGGGGGGTGGCGAAGAGACCAGAGACTATGACATCACCAGCTCCACGA	2213
Q	y	721	GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro	740
D	b	2214	GGTCTGGAGGCAGCGCCGGAGTGGTTCTCCGCAATGACGTGGGCACCAACCATCATCCG	2273
Q	y	741	ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu	760
D	b	2274	ACACCATGTACCTCTCTAGCCAGCCAACCCAGATGAATCGSAACTTTATAATTGAG	2333
Q	y	761	AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe	780
D	b	2334	AACCTGAAGCGGGTAAACACAGACCCACAGCCCCGCCCTACGACACCCCTTTGGTGTTC	2393
Q	y	781	AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer	800
D	b	2394	GACTATGAGGGCASC GGCTCCGAGCGCGCTCCCTGAGCTCCCTACCTCCTCGGCTCC	2453
Q	y	801	AspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGlySerArgPheLysLysLeuAla	820
D	b	2454	GACCAAGACCAAGATTACGATTATCTGAACAGTGGGGCGACCGCTTCAAGAGCTGGCA	2513
Q	y	821	AspMetTyrGlyGlyGlyGluAspAsp	829
D	b	2514	GACATGTACGGTGGCGGGAGGACGAC	2540
R	E	S	RESULT 3	
L	O	C	AX829232	
D	E	F	LOCUS	
D	E	F	Sequence 125 from Patent WO02059377.	
D	E	F	AX829232	
D	E	F	Accession AX829232.1 GI:39838965	
D	E	F	Keywords Homo sapiens (human)	
D	E	F	Source Homo sapiens	
D	E	F	ORGANISM Homo sapiens	
D	E	F	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
D	E	F	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
D	E	F	1	
D	E	F	Authors Mack D.H., Gish K.C. and Afar,D.	
D	E	F	Title Methods of diagnosis of breast cancer, compositions and methods of screening for modulators of breast cancer	
D	E	F	Journal Patent: WO 02059377-A 125 01-AUG-2002;	
D	E	F	EOS Biotechnology, Inc. (US)	
D	E	F	Features Location/Qualifiers	
D	E	F	source	
D	E	F	1..3171	
D	E	F	/organism="Homo sapiens"	
D	E	F	/mol_type="unassigned DNA"	
D	E	F	/db_xref="taxon:9606"	
D	E	F	ORIGIN	
D	E	F	Alignment Scores:	
D	E	F	Pred. No.: 0 Length: 3171	
D	E	F	Score: 829.00 Matches: 829	
D	E	F	Percent Similarity: 100.00% Conservative: 0	
D	E	F	Best Local Similarity: 100.00% Mismatches: 0	
D	E	F	Query Match: 100.00% Indels: 0	
D	E	F	DB: 6 Gaps: 0	
D	E	F	US-09-916-849A-1 (1-829) x AX829232 (1-3171)	
Q	y	1	MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuGlnValCystTrpLeuGln	20
D	b	54	ATGGGGCTCCCTCGTGGACCTCTCGGTCTCTCCCTCTCCAGTGTGGTGGGTGAG	113
Q	y	21	CysAlaAlaSerGluProCvsArqAlaValPheArgGluAlaGluValThrLeuGluAla	40

114 TCGCGGCTCCGAGCGTCCGCGGCTTCAGGAGGCTGAAGTACCTTGAGGCG 173
41 GlyAlaGluGluProGlyGlnAlaLeuGlyValPheMetGlyCysProGly 60
174 CGAGCGCGGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 233
61 GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr 80
234 CAAAGACGAGCTCTGTAGCACTGATATGATGACTTCACTGTGCGGAATGGCGAGACA 293
81 ValGlnGluArgSerLeuLeuGluArgAsnProLeuLeuValPheProSerLeuArg 100
294 GTCCAGGAAGAGGTCACTGAAGGAAGGAATCCATTTGAAGATCTCCCATCCAAACGT 353
101 IleLeuArgArgHisLeuArgAspTrpValAlaAlaProLieserValProGluAsnGly 120
354 ATCTTACGAGACACAGAGAGATTGGGTGTCTCCATATCTGTCCCTGAAATGGC 413
121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
414 AAGGGTCCCTCCCGCAGAGATGAATCAGCTCAAGTCTAATAAAGATAGACACCAAG 473
141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProProGluGlyValPheAlaVal 160
474 ATTTTCTACAGCATCACGGGGCGGGGCGAGACGCCCTTGAGGGTCTTTCGCTGTA 533
161 GluLysGluThrGlyTrpLeuLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys 180
534 GAGAGGACAGAGCTGGTGTGTGTGAATAAGCCACTGGACCGGAGAGATTCCCAAG 593
181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
594 TATGAGCTTTTGGCCACGCTGTGTCAGAGAAATGGTGTCTCAGTGAGAGACCCCATGAAC 653
201 IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
654 ATCTCCATCATGTGACCGACCAAGATGACCAAGCCCAAGTTTACCAGAGACCTTC 713
221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
714 CGAGGAGGTCTTATAGAGGAGTCTTACCAGGTACTTCTGTGATGACGAGTACAGCCACA 773
241 AspGluAspAspAlaIleThrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln 260
774 GATGAGATGATGCCATCTACACCTACAAATGGGTGGTTCCTTACTCCATCCATAGCCAA 833
261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
834 GAACCAAGACCCACACGACCTCATGTTCAAAATTCACCGAGACACGACCATCAGC 893
281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
894 GTCATCTCCAGTGGCTGGACCGGGAAGTCCCTGAGTACACACTGACCATCCAGGCC 953
301 ThrAspMetAspGlyAspGlySerThrThrThrAlaValAlaValValGluIleLeuAsp 320
954 ACAGACATGGATGGGACGGCTCCACCACCGCAGTGCAGTAGTAGGATCTTGAT 1013
321 AlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
1014 GCCAATGACAATGCTCCCTGTTTACCCCGCAGAGTACGAGGCCCATGTGCTCAGAAAT 1073
341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
1074 GCAGTGGGCCATGAGGTGCAGAGCTGACGGTCACTGATCTGACGCCCGCCCACTCACA 1133
361 AlaTrpArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
1134 CGGTGGCTGCCACCTTACCTTATCATGGCGGTGACGACGGGAGCCATTTTACCATCACC 1193
381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400

1194 ACCACCTCGAGAGCAACACGAGCGCATCTGACAAACAGGAGGGTTTGGATTTTGGCGC 1253
401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
1254 AAAAACCGACGACACCTCTGTACGTTGAAGTGACCAACGAGGCGCTTTTGTGCTGAAGCTC 1313
421 ProThrSerThrAlaThrIleValHisValGluAspValAsnGluAlaProValPhe 440
1314 CCAACCTCCACAGCCACCATAGTGGTCCACGCTGGAGGATGTGAATGAGGACCTGTGT 1373
441 ValProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
1374 GTCCACCTCCAAAGTCGTTGAGGTCAGAGGGCATCCCCACTGGGAGCCTGTGTGT 1433
461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
1434 GTCTACACTGCAGAGACCCCTGACAAAGGAAATCAAAAGATCAGTACCACCTCTGAGA 1493
481 AspProIleGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
1494 GACCCACGAGGGTGGCTAGCCATGGACCCAGACAGTGGGAGGTACAGCTGGGGCACC 1553
501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
1554 CTCGACCTGAGGATGAGCAGTTTGTGAGGAACAACATCTATGAAGTCATGGTCTTGCC 1613
521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAsp 540
1614 ATGACAAATGGAAGCCCTCCACACTGGCAGCGGAACCCCTCTGTCTAACTGATGAT 1673
541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
1674 GTCACCAACCATGGCCCGGCTCCCTGAGCCCGCTCAGATCACCATCTGCACCAAGCCCT 1733
561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
1734 GTGGCCACGCTGTGAACATCAGGACCAAGGACCTGTCTCCACACCTCCCTTTCAG 1793
581 AlaGlnLeuThrAspAspSerAspIleTyrThrAlaGluValAsnGluGluGlyAsp 600
1794 GCCCAGCTCAGATGACTCAGACATCTACTGACCGCAGAGGTCAACGAGGAGGTGAC 1853
601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
1854 ACAGTGTCTTGTCCCTGAAGAGTTCCTGAAGCAGGATACATATGACGTGCACCTTCT 1913
621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
1914 CTGTCTGACCATGGCAACAAAGACAGCTCAGCGTGTATCAGGGCCACTGTGTGCGACTGC 1973
641 HisGlyHisValGluThrCysProGlyProTriplysGlyGlyPheIleLeuProValLeu 660
1974 CATGGCCATGTCGAAACCTGCCCCGAGCCCTGGAAAGGAGGTTTCATCCTCCCTGTGCTG 2033
661 GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysLys 680
2034 GGGGCTGTCTGGCTGTCTGTCTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2093
681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
2094 CGGAAGATCAAGAGACCCCTCTACTCCGAGAAGATGACCCCGTGACACGCTTCTTCTAC 2153
701 TyrGlyGluGluGlyGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
2154 TATGGCAAGAGGGGGTGGCGAAGAGGACCCAGGACTATGACATCACCAGCTCCACCGA 2213
721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleLeuPro 740
2214 GGTCTGAGGCGCAGCGCGGAGGTGTCTCGCAATACGCTGGGACCAACCATCATCCCG 2273
741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleGlu 760
2274 ACACCATGTACCTCTCTAGCCAGCCAAACCCACATGAATCGCAACTTTATAATTGAG 2333

Qy	761	AsnLeuIysAlaIasnThrAspProThrAlaProProFtyrAspThrLeuLeuValPhe	780
Db	2334	AACTGAAGCGCGCTAACACAGACCCACAGCGCCCGCTTAGCACCCCTCTGTGTGTC	2393
Qy	781	AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer	800
Db	2394	GACTATGAGCGCAGCGCTCCGACGCGCGTCCCTGAGCTCCCTACCTCCCTCGCGCTCC	2453
Qy	801	AspGlnAspGlnAspTyrAspTyrLeuAsnGluTyrGlySerArgPheIysLysLeuAla	820
Db	2454	GACCAAGACCAAGATTACGATTATCTGAACGAGTGGGCGAGCGCTTCAAGAGCTGGCA	2513
Qy	821	AspMetTyrGlyGlyGluAspAsp	829
Db	2514	GACATGTACGTCGCGGAGGAGCAGAC	2540
RESULT 4			
LOCUS	HSPCAD	3171 bp	mRNA linear PRI 06-FEB-1992
DEFINITION	H sapiens mRNA for p cadherin.		
ACCESSION	X63629		
VERSION	X63629.1	GI:35322	
KEYWORDS	cadherin.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. Shimoyama, Y., Yoshida, T., Terada, M., Shimosato, Y., Abe, O. and Hirohashi, S. Molecular cloning of a human Ca ²⁺ -dependent cell-cell adhesion molecule homologous to mouse placental cadherin: its low expression in human placental tissues J. Cell Biol. 109 (4 Pt 1), 1787-1794 (1989)		
JOURNAL	90009051		
MEDLINE	2793940		
FEATURES	Location/Qualifiers		
source	1..3171		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
CDS	54..2543		
	/codon_start=1		
	/product="p-cadherin"		
	/protein_id="CAA45177.1"		
	/db_xref="GI:35323"		
	/db_xref="GOA:P22223"		
	/translation="MGLPRGLASILLQVHLCASSEPCRAVFEAEVLEAGAE QEPQALGVKFGCPQGEPALESTNDQFTVRNGSTQVERSLKERNPKLIPPSKRIL RRKQDWVAPISVPENGKPPQRLNQLKSNKRDTKI FYSTPGADSPPEGVAV EKETGWLKLPDLREIEIAKYLFGHAYSENGASVEDPMNISII VTDQNDHKPKFTQD TFRGSVLGVLPGTSMQVATDEDDAIYTVNGVAYSIIHSEPKDHLDMFTIHRST GTIVSSGLDRKPYEYTLTIQATMDGDGSTTTAVAVETLIDANDNAPWFDPOKYE AHVPNAGVHQRLVTLDLADPNPAMRATYLLINGDDGDHFTITHPESNQGLIT RKLGFPAQKHLIVETNEAPFKLPTSTATIVHEVDNEAPVFPVSKVEVQ EGIPTEPVCVTAEDPKENOKISYRILRDPAGLMDPDSQQTAVGTDLRDEQF VARNIYEVMLVANDNGSPPTGTLLTLIDVNDHGPVPEPQITICNQSPVRHLN ITDKOLSPTSFQALTDSDSIYWTAVNEEGDVTVLKXKFLQDITYDHLSLSH GNKEQLTVTRATVCHGVETCPGWKGFILPVLGVALLLFLLVLLLVLRKRRK IKEPLLIPDDTRDNVYVYEGSGGEEQDYDITOLHGLEAPPEWLRNDVAPTIIP TWVPRPANPEIGNFIENLKNATPTAPPTLTLVFDYBGSQSDAASLSLTS ASQDQDYDLNWSRFAKLADNYGGEDD"		
polyA_signal	3162..3167		
ORIGIN			
Alignment Scores:			
Pred. No.:	0	Length:	3171
Score:	829.00	Matches:	829
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0

DB:	9	Gaps:	0
US-09-916-849A-1 (1-829) x HSPCAD (1-3171)			
Qy	1	MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuGlnValCystTyrLeuGln	20
Db	54	ATGGGGCTCCCTCGTGACCTCTCGCGTCTCTCTCTCTCTCCAGGTTTCTGGCTGCAG	113
Qy	21	CysAlaIaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla	40
Db	114	TGCGCGCCCTCCGAGCGCTCCGCGCGTCTTCCAGGAGCTCAAGTGACCTTGGAGGG	173
Qy	41	GlyGlyAlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly	60
Db	174	GGAGCGCGGAGGAGGAGCCCGCGCGCTGGGNAAGTATTCATGGCTGCCCTGGG	233
Qy	61	GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr	80
Db	234	CAAGAGCCAGCTCTGTTTAGCACTGATAATGATGACTTCACTGTGGGAATGCGGACA	293
Qy	81	ValGlnGluArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg	100
Db	294	GTCAGGAAAGAGGTCACTGAAGAAAGGAATCCATTGAAGATCTTCCATCCAAAGT	353
Qy	101	IleLeuArgArgHisLysArgAspTyrValValAlaProIleSerValProGluAsnGly	120
Db	354	ATCTTACGAAGACACAAAGAGAGATTGGGTGGTTCCTCAATATCTGCTCCCTGAAATGGC	413
Qy	121	LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgThrLys	140
Db	414	AAGGTCCTTCCCTCCAGACACTGAATCAGTCAAGTCTAATAAAGATAGAGACCAAG	473
Qy	141	IlePheTyrSerIleThrGlyProGlyAlaAspSerProGluGlyValPheAlaVal	160
Db	474	ATTTTCTACAGCATCAGCGGGCGGGGAGACAGCCCCCTCAGGGGTGTCTTCGTGTA	533
Qy	161	GluLysGluThrGlyTyrLeuLeuLeuAsnLysProLeuAspArgGlnGluLeuLys	180
Db	534	GAGAAGGAGACAGCGCTGTTGTTGTAATAAGCCACTGACCGGAGGAGATTGCCAAG	593
Qy	181	TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn	200
Db	594	TATGAGCTCTTTGGCCAGCGTGTGTGAGAAATGGTCCCTCAGTGGAGACCCCATGAC	553
Qy	201	IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe	220
Db	654	ATCTCCATCATCTGACCGACCAAGATGACCAAGCCCAAGTTTACCCAGGACACCTTC	713
Qy	221	ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr	240
Db	714	CGAGGGAGTGCTTTAGAGGAGTCCTACAGGTACTTCTGTGTGATGAGGTGACAGCCACA	773
Qy	241	AspGluAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln	260
Db	774	GATGAGGATGATGCCATCTACACTCAATGGGGTGGTGTCTTACTCCATCCATAGCCAA	833
Qy	261	GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer	280
Db	834	GACCAAGAGACCCACACGACCTCATGTTTCAATTCACCGGAGCAGACGACCATCAGC	893
Qy	281	ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla	300
Db	894	GTCTCTCCAGTGGCTGGACCGGGAAGTCCCTGAGTACACACTGACCATCCAGGCC	953
Qy	301	ThrAspMetAspGlyaspGlySerThrThrAlaValAlaValValGluLeuLeuAsp	320
Db	954	ACAGACATGGATGGGACGCGCTCCACCCACCGGAGTGGAGTAGTGAGATCTTGAT	1013
Qy	321	AlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn	340
Db	1014	GCCAAATGACATGCTCCCATGTTTACCCCCAGAGTACAGGCCCATCTGCTGTAGAT	1073
Qy	341	AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro	360

Db 1074 GCAGTGGGCCATGAGGTGCAGAGGCTGACGGTCACTGATCTGAGCCGCCCACTCACCA 1133
 QY 361 AlaTrpArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
 Db 1134 GGTGGCTGCTCCACCTTATCATGGGGGGTGCAGCGGACCACTTTTACCATCAC 1193
 QY 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgGlyGlyLeuAspPheGluAla 400
 Db 1194 ACCACCCCTGAGAGCAACAGGGGCTCTGACAAACAGGAGGGTTTGGATTTTGGAGCC 1253
 QY 401 LysAsnGlnHisPheThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
 Db 1254 AAAAACAGACACACCTGTACGTGAGGTGACCAACAGGGGCCCTTTTGTGTGAGGCTC 1313
 QY 421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
 Db 1314 CCAACCTCCACAGCCACCATAGTGTCCAGCTGGAGGATGTGAATGAGGCACCTGTGTT 1373
 QY 441 ValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
 Db 1374 GTCCCAACCTCCAAAGTCGTGTGAGGTCCAGGAGGCGATCCCACTGGGAGGCTGTGTG 1433
 QY 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
 Db 1434 GTCTACACTGCAGAGACCTGACAGGAGATCAAAAGATCAGCTACCGCATCTGAGA 1493
 QY 481 AspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
 Db 1494 GACCCAGCAGGCTGGCTAGCATGAGCAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1553
 QY 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
 Db 1554 CTCGACCGTGGAGATGAGCAGTTGTGAGGAGCAACATCATGAAGTCATGCTCTGTGCC 1613
 QY 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAsp 540
 Db 1614 ATGGCAATGGAAGCCCTCCACCACTGCGACCGGAACCTCTGCTAACACTGATTGAT 1673
 QY 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
 Db 1674 GTCAAGCAGCATGGCCAGTCCCTGAGCCCTGATGATCACCATCTGCAACCAAGCCCT 1733
 QY 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
 Db 1734 GTGGGCCAGCTGCTGACATCAGGACAGGACCTGTCTCCCAACCTCCCTTCCAG 1793
 QY 581 AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGlyAsp 600
 Db 1794 GCCCAGCTCACAGATGACTCAGACATCTACTGACGGCAGAGGTCAACGAGGAGGTGAC 1853
 QY 601 ThrValValLeuSerLeuLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
 Db 1854 ACAGTGGCTTGTCCCTGAAGAAGTTCTGAAGCAGGATACATATGAGCTGACCTTTCT 1913
 QY 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
 Db 1914 CTGTCTGACATGGCAACAAAGAGCAGTGCAGGTGATCAGGGCCACTGTGTGCGACTGC 1973
 QY 641 HisGlyHisValGluThrCysProGlyProThrLysGlyGlyPheIleLeuProValLeu 660
 Db 1974 CATGGCCATGTGAACCTGCTGACCTGACCTGGAAGGAGGTTCATCTCCCTGCTGTG 2033
 QY 661 GlyAlaValLeuAlaLeuLeuPheLeuLeuLeuValLeuLeuValArgLysLys 680
 Db 2034 GGGGCTGTCTGCTGTCTGTCTCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2093
 QY 681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
 Db 2094 CGAAAGATCAAGAGGCCCTCTCTACTCCAGAGATGACACCCGTGACAAAGCTCTCTAC 2153
 QY 701 TyrGlyGluGluGlyGlyGluGluAspGlnAspTyrAspIleThrGluLeuHisArg 720

Db 2154 TATGCCGAAGAGGGGGTGGCGAAGAGGACCAGGACTATGACATCACCCAGCTCCACGA 2213
 QY 721 GlyLeuGluAlaAspProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
 Db 2214 GGTCTGAGGGCCAGCGGAGGTGTCTTCCGCAATACGCTGGCACCACCATCATCCG 2273
 QY 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
 Db 2274 ACACCATGTACCTCTAGCCAGCAACCCAGATGAATCGCAACTTATTAATTGAG 2333
 QY 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuValPhe 780
 Db 2334 AACCTGAAGGGGCTAACACAGACCCACAGCCCGCCCTACGACACCTCTTGGTGTTC 2393
 QY 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
 Db 2394 GACATGAGGGCAGCGCTCCGACGCGCTCCCTGAGCTCCCTCACTCTCCGCTCC 2453
 QY 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGlySerArgPheLysLeuAla 820
 Db 2454 GACCAAGACCAAGATTACGATTATCTGACGAGTGGGCGAGCCGCTTCAAGAAGCTGCA 2513
 QY 821 AspMetTyrGlyGlyGlyGluAspAsp 829
 Db 2514 GACATGTACGTGGCGGGAGGAGCAG 2540

RESULT 5
 AX382118 3219 bp DNA linear PAT 18-MAR-2002
 LOCUS AX382118
 DEFINITION Sequence 17 from Patent WO0200939.
 ACCESSION AX382118
 VERSION AX382118.1 GI:19576927
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE
 1 Macina, R.A. and Pillai, R.
 METHOD of diagnosing, monitoring, staging, imaging and treating
 colon cancer
 PATENT: WO 0200939-A 17 03-JAN-2002;
 JOURNAL Diadexis, Inc. (US)
 FEATURES
 source 1..3219
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"

ORIGIN
 Alignment Scores:
 Pred. No.: 0 Length: 3219
 Score: 715.00 Matches: 828
 Percent Similarity: 99.64% Conservative: 0
 Best Local Similarity: 99.64% Mismatches: 1
 Query Match: 86.25% Indels: 3
 DB: 6 Gaps: 0

US-09-916-849A-1 (1-829) x AX382118 (1-3219)
 QY 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuGlnValCysTrpLeuGln 20
 Db 90 ATGGGGCTCCCTCGTGGACCTCTCGGCTCTCTCTCTCTCCAGGTTTCTGGCTGCAG 149
 QY 21 CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
 Db 150 TGGCGGCTCCCGAGCGCTGCGGCGGTCTTTCAGGAGGCTGAAGTACCTTGGAGGCG 209
 QY 41 GlyGlyAlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
 Db 210 GGAGCGCGGAGCAGGAGCCCGGCGAGCGCTGGGAAAGATTTCATGGGCTGCCCTGGG 269
 QY 61 GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr 80

Db 270 CAAGAGCCAGCTCTGTTTAGCCTAGTAATGATGACTTCTACCTGTGCGAATGGCAGACA 329
Qy 81 ValGlnGluArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
Db 330 GTCCACGAAAGAGGTCACCTGAAGGAAAGAAATCCATTGAAGATCTTCCCATCAAACGT 389
Qy 101 IleLeuArgHisLysArgAspTrpValAlaProIleSerValProLusnGly 120
Db 390 ATCTTACGAAAGACACAAAGAGATGGGTGGTGTCTCCAAATATCTGTCCCTGAAATGGC 449
Qy 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
Db 450 AAGGTCCTTCCCCCAGAGACTGAATCAGCTCAAGTCTAATAAGATAGACACCAAG 509
Qy 141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProProGluGlyValPheAlaVal 160
Db 510 ATTTTCTACAGCATCACGGGGCGGGGCGAGACAGCCCCCTGAGGGTGTCTTCCCTGTA 569
Qy 161 GluLysGluThrGlyTrpLeuLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys 180
Db 570 GAGAGAGAGACAGGCTGGTGTGTTGTAATAGCCACTGGACCGGGAGAGATGCCAAG 629
Qy 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
Db 630 TATGAGCTCTTTGGCCACCGCTGTGCAGAGAAATGGTGCCTCAGTGGAGAGACCCCATGAAC 689
Qy 201 IleSerIleLeuValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
Db 690 ATCTCCATCATCGTACCCAGACAGATGACCAAGCCCAAGTGTACCCAGGACACCTTC 749
Qy 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
Db 750 CGAGGGAGTGTCTTAGGGGAGTCTTACCAGGTACTTCTGTGATGCAGGTGACAGCCACA 809
Qy 241 AspGluAspAspAlaIleTyrThrThrAsnGlyValValAlaTyrSerIleHisSerGln 260
Db 810 GATGAGGATGATCCCATCTACACCTACAAATGGGGTGTGTCTTACTCCATCCATAGCCAA 869
Qy 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
Db 870 GAACCAAGAGACCCACACGACCTCATGTTCAAAATTCACCGGAGCACAGGCACCATCAGC 929
Qy 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
Db 930 GTCATCTCCAGTGGCCTGGACCGGGAAAAAGTCCCTGAGTACACACTGACCATCCAGGCC 989
Qy 301 ThrAspMetAspGlyAspGlySerThrThrAlaValAlaValValGluIleLeuAsp 320
Db 990 ACAGATGATGGGACCGCTCCACACACGCGAGTGGCAGTAGTGAGATCTTGTAT 1049
Qy 321 AlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
Db 1050 GCCAATGACAATGCTCCCATGTTTACCCGCCAGAGTACGAGGCCCATGTGCTGAGAAAT 1109
Qy 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
Db 1110 GCAGTGGGCGCATGAGGTGACAGGCTGAGGCTCACTGATCTGAGCGCCCCCACTACCA 1169
Qy 361 AlaTrpArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
Db 1170 GCGTGGGCTGCCACCTACCTTATCATGGCGGTGACGACGGGACCATTTTACCATCACC 1229
Qy 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
Db 1230 ACCCACCTCAGAGACCAACCGGCATCTTGCAGAACCCAGAGGGTGTGATTTGAGGCC 1289
Qy 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db 1290 AAAAAACGACACACCTGTGTACGTTGAAGTACCAACGAGGCCCTTTTGTGTGAAGCTC 1349
Qy 421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440

Db 1350 CCAACCTCCACAGCCACCACATAGTGGTCCACGTGAGGATGTGAATGAGGCACCTGTGTTT 1409
Qy 441 ValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
Db 1410 GTCCACCTCCAAAGTCGTGTAGGTCCAGAGGGCATCCCCACTGGGAGCCTGTGTGT 1469
Qy 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db 1470 GTCTACACTCGAGAGACCCTGACAGGAGAATCAAAAGATCAGCTACCGCATCTGAGA 1529
Qy 481 AspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db 1530 GACCCAGCAGGGTGGCTAGCCATGGACCCACAGAGTGGGAGGTCAACGCTGTGGGCACC 1589
Qy 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db 1590 CTGACCGCTGAGGATGAGCAGTTTGTGAGGAACAACATCTATGAGTCTGCTTTGGCC 1649
Qy 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAsp 540
Db 1650 ATGGACAATGGAAGCCTCCACCACTGGCAGCGGAACCCCTCTGTCTAACACTGATTGAT 1709
Qy 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 1710 GTCAACGACCATGGCCCGCAGTCCCTGAGCCCCCTGAGATCAACATCTGCAACCAAGCCCT 1769
Qy 561 ValArgHis-Val-LeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheG 580
Db 1770 GTGCGCCA-GTGGCTGAACATCACGGAACAAGACCTGTCTCCCCACACCTCCCTTTCC 1828
Qy 580 InAlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGluGlyA 600
Db 1829 AGGCCACAGCTCACAGATGACTCAGACATCTACTGGACCGCAGAGGTCAACAGAGAAGGTG 1888
Qy 600 spThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuS 620
Db 1889 ACACAGTGGTCTTGTCTCCGAGAGATCTCTGAGCAGGATACATATGACGTGCACCTTT 1948
Qy 620 erLeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspC 640
Db 1949 CTCTGTCTGACCATGGCAACAAGAGCAGCTGACGCTGATCAGGCGCACCTGTGTGGACT 2008
Qy 640 YHisGlyHisValGluThrCysProGlyProTrpLysGlyGlyPheIleLeuProValL 660
Db 2009 GCCATGGCCATGTCGAAACCTGCCCTGGACCTTGGAGGGAGGTTTCATCTCCTGTGC 2068
Qy 660 euGlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysL 680
Db 2069 TGGGGGCTGTCTTGGCTCTGCTCTCTCTCTGCTGTCTTGTGTGGTGGTGGTGGTGGT 2128
Qy 680 YSArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheT 700
Db 2129 AGCGGAAGATCAAGAGGCCCTCTTACTCCAGAGATGACACCCCGTGAACAGCTTCTCT 2188
Qy 700 YrTyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisA 720
Db 2189 ACTATGCGGAAGAGGGGGTGGCGAGAGGACAGGACTATGACATCACCCAGCTCCACC 2248
Qy 720 rgGlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIleP 740
Db 2249 GAGTCTCTGGAGGCCAGGCCGGAGTGGTCTTCCGCAATACGTGGGACCAACCATATCC 2308
Qy 740 roThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleG 760
Db 2309 CGACACCATGATCCGCTCTTACGCCAGCAACCCAGATGAATCGCAACTTTATAATTG 2368
Qy 760 luAsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValP 780
Db 2369 AGAACCTGAAGGGGGCTTAACACAGACCCACAGCCCCCGCCCTACGACACCTCTTGGTGT 2428
Qy 780 heAspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaS 800
Db 2429 TCGACTATGAGGGCAGGGGCTCCGACCGCCGCTCCCTGAGCTCCCTCACCTCTCCGCTC 2488

Db 2943 CCGACCAACACAGATACAGTATCTGACGAGTGGGCGAGCGCTTCAGAGCTGG 3002
QY 820 laAspMetTyrglyGlyGluAsp 829
Db 3003 CAGACATGACGTTGGCGGGGAGGACGAC 3031
RESULT 7
AX210689 306 bp DNA linear PAT 31-AUG-2001
LOCUS
DEFINITION Sequence 331 from Patent WO0157058.
ACCESSION AX210689
VERSION AX210689.1 GI:15424949
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE Detection of differential gene expression
JOURNAL Patent: WO 0157058-A 331 03-AUG-2001;
Metagen Gesellschaft fuer Genomforschung mbH (DE)
FEATURES
source
1. 306
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 3.46e-98 Length: 306
Score: 102.00 Matches: 102
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 12.30% Indels: 0
DB: 6 Gaps: 0
US-09-916-849A-1 (1-829) x AX210689 (1-306)
QY 707 GlyGluGluAspGlnAspTyraAspThrGlnLeuHisArgGlyLeuGluAlaArgPro 726
Db 1 GCGAGAGAGGACGAGGACTATGATCACCACCCAGCTCCACCGAGGCTGGAGGCCAGCGG 60
QY 727 GluValValLeuArgAsnAspValAlaProThrIleleProThrProMetTyraArgPro 746
Db 61 GAGGTGGTTCTCCGCAATGACGTGGCACCAACATCATCCCGACACCCATGTACCGTCT 120
QY 747 ArgProAlaAsnProAspGluIleGlyAsnPhelIleGluAsnLeuIysAlaAlaAsn 766
Db 121 CGGCCAGCCCAACCCAGATGAATCGCAACTTTATTAATTGAACTGAGGCGGCTTAC 180
QY 767 ThrAspProThrAlaProProTyraAspThrLeuLeuValPheAspTyrglySerGly 786
Db 181 ACAGACCCACAGCCCGCCCTACGACACCTCTTGGTGTTCGACTATGAGGCGGCGGC 240
QY 787 SerAspAlaSerLeuSerSerLeuThrSerSerAlaSerAspGlnAspGlnAspTyra 806
Db 241 TCGACGCGCGGCTCCCTGAGCTCCCTCACCTCTCCGCTCCGACCAACCAAGATTAC 300
QY 807 AspTyr 808
Db 301 GATTAT 306
RESULT 8
AC099314 184945 bp DNA linear PRI 05-APR-2003
LOCUS
DEFINITION Homo sapiens chromosome 16 clone RP11-354M1, complete sequence.
ACCESSION AC099314
VERSION AC099314.3 GI:29568028
KEYWORDS
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Direct Submission
JOURNAL
REFERENCE 2 (bases 1 to 184945)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL
REFERENCE 3 (bases 1 to 184945)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL
REFERENCE 4 (bases 1 to 184945)
AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
TITLE Submitted (05-APR-2003) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
JOURNAL
COMMENT On Apr 5, 2003 this sequence version replaced gi:18057085.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center and Los Alamos
National Laboratory
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.8% of Sequence;
Estimated Total Number of Errors is 0.4.
FEATURES
source
1. 184945
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-354M1"
ORIGIN
Alignment Scores:
Pred. No.: 3.07e-72 Length: 184945
Score: 80.00 Matches: 80
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.65% Indels: 0
DB: 9 Gaps: 0
US-09-916-849A-1 (1-829) x AC099314 (1-184945)
QY 395 GlyLeuAspPheGluAlaLysAsnGlnHisThrLeuTyraValGluValThrAsnGluAla 414
Db 28107 GGTTTGGATTTTGAGGCCAAACACACACACCTGTACCTGTACCTGACCTACGAGGCC 28166
QY 415 ProPheValLeuLysLeuProThrSerThrAlaThrIleValHisValGluAspVal 434
Db 28167 CCTTTTGCTGAGAGCTCCCAACCTCCACAGCCACCATAGTGTCCACGTGGAGGATGTG 28226
QY 435 AsnGluAlaProValPheValProProSerLysValValGluValGlnGluGlyIlePro 454
Db 28227 AATGAGGCACCTGTGTTGTGTCCTCCACCTCCAAAGTCCTTGGGTCACAGGAGGATCCCC 28286
QY 455 ThrGlyGluProValCysValTyThrAlaGluAspProAspLysGluAsnGlnLysIle 474
Db 28287 ACTGGGGAGCCTGTGTGTGTCTACCTGCAGAGACCTCTGACAGGAGAGATCAAGATC 28346
RESULT 9
CNS01DXD/c
ID CNS01DXD standard; genomic DNA; HTG; 186939 BP.
XX
AC AL139227;
XX

SV AL139227.2
 XX 11-FEB-2000 (Rel. 62, Created)
 DT 25-MAY-2000 (Rel. 63, Last updated, Version 2)
 XX
 DE Human chromosome 16 DNA sequence *** IN PROGRESS *** BAC R-61512 of library
 DE RPII-11 from chromosome 16 of Homo sapiens (Human)
 XX
 KW HTG; HTGS_DRAFT; HTGS_PHASE2.
 XX
 XX Homo sapiens (human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
 XX
 RN [1]
 RP 1-186839
 RA Genoscope;
 RT Submitted (23-MAY-2000) to the EMBL/GenBank/DBJ databases.
 RL
 XX

IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage, etc. . even if efforts are made to eliminate these contaminating sequences. The following BAC sequence is oriented from the T7 to the SP6 end.
 CC Contigs composition :
 CC 49057 bp contig from 1 to 49057
 CC 31448 bp contig from 49158 to 80605
 CC 55190 bp contig from 80706 to 135895
 CC 3703 bp contig from 135996 to 139698
 CC 5159 bp contig from 139799 to 144957
 CC 3046 bp contig from 145058 to 148103
 CC 14904 bp contig from 148204 to 163107
 CC 23632 bp contig from 163208 to 186839
 XX

Overall quality chart :
 CC Range : bases
 CC 0 : 1083
 CC 1 - 9 : 1569
 CC 10 - 19 : 2752
 CC 20 - 29 : 6935
 CC 30 - 39 : 18862
 CC 40 - 49 : 16913
 CC 50 - 59 : 20166
 CC 60 - 69 : 35143
 CC 70 - 79 : 45986
 CC 80 - 89 : 26801
 CC 90 - 99 : 10629
 CC

Percentage of bases with a quality value >= 40 : 83 %
 CC
 XX Key Location/Qualifiers
 FH source
 FT 1. 186839
 FT /chromosome="16"
 FT /db_xref="taxon:9606"
 FT /mol_type="genomic DNA"
 FT /organism="Homo sapiens"
 FT /clone_lib="RPII-11"
 FT /clone="R-61512"
 XX

Sequence 186839 BP; 49577 A; 44657 C; 43147 G; 48728 T; 730 other;
 Alignment Scores:
 Pred. No.: 3,1e-72 Length: 186839
 Score: 80.00 Matches: 80
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 9,65% Indels: 0
 DB: 30 Gaps: 0

US-09-916-849a-1 (1-829) * CNS01DXD (1-186839)
 QY 395 GlyLeuAspPheGluAlaLysAsnGlnHisThrLeuTyrValGluValThrAsnGluAla 414
 DB 18582 GGTTCGATTTCAGGCCAAACACACACACCTGACCTGTGAGTGACCAACGAGGCC 18523
 QY 415 ProPheValLeuLysLeuProThrSerThrAlaThrIleValHisValGluAspVal 434
 DB 18522 CCTTTGTGCTGAAGCTCCCAACCTCCACACCCACCATAGTGTGTCACGTCGAGGATGTG 18463
 QY 435 AsnGluAlaProValPheValProProSerLysValValGluValGluGlyIlePro 454
 DB 18462 AATGAGGCACCTGTGTTGTCTCCACCTCCAAAGTCGTTGAGGTCAGGAGGCATCCCC 18403
 QY 455 ThrGlyGluProValCysValTyrThrAlaGluAspProAspLysGluAsnGlnLysIle 474
 DB 18402 ACTGGGAGCCTGTGTGTCTACTGTCAGACAGACCTGACAAGGAGATCAAAAGATC 18343
 RESULT 10
 AC126773 196911 bp DNA linear PRI 26-NOV-2003
 LOCUS Homo sapiens chromosome 16 clone RPII-61512, complete sequence.
 DEFINITION AC126773 AL139227
 ACCESSION AC126773.3 GI:38524628
 VERSION AC126773.3
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 196911)
 AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los Alamos National Laboratory.
 TITLE Direct Submission
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 196911)
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (09-JUL-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 REFERENCE 3 (bases 1 to 196911)
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (23-DEC-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 REFERENCE 4 (bases 1 to 196911)
 AUTHORS Stanford Human Genome Center and Los Alamos National Laboratory.
 TITLE Direct Submission
 JOURNAL Submitted (26-NOV-2003) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 COMMENT On or before Nov 26, 2003 this sequence version replaced GI:9212225, gi:27363207.
 Draft Sequence Produced by DOE Joint Genome Institute
 www.jgi.doe.gov
 Finishing Completed at Stanford Human Genome Center and Los Alamos National Laboratory
 www.sngc.stanford.edu
 Quality: Phrap Quality >=40 100% of Sequence;
 Estimated Total Number of Errors is 0.
 Location/Qualifiers
 1. 196911
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="16"
 /clone="RPII-61512"
 FEATURES
 source
 ORIGIN
 Alignment Scores:
 Pred. No.: 3,25e-72 Length: 196911
 Score: 80.00 Matches: 80
 Percent Similarity: 100.00% Conservative: 0

```

Best Local Similarity: 100.00%      Mismatches: 0
Query Match:          9.65%        Indels: 0
DB:                   9            Gaps: 0

US-09-916-849A-1 (1-829) x AC126773 (1-196911)

QY 395 GlyLeuAspPheGluAlaLysAsnGlnHisThrLeuYrValGluValThrAsnGluAla 414
DB 178220 GATTTGGATTTGAGGCCAAACCAACGACACCCCTGTACGTTGAAGTACCAACGAGGCC 178279
QY 415 ProPheValLeuLysLeuProThrSerThrAlaThrIleValValHisValGluAspVal 434
DB 178280 CCTTTTGTCTGAAGCTCCCAACCTCCACAGCCACCATAGTGGTCCACGTCGAGGATGTG 178339
QY 435 AsnGluAlaProValPheValProProSerLysValValGluValGlnGluGlyIlePro 454
DB 178340 AATGAGGACACCTGTGTTGTCCACCCCTCCAAAGTCGTTGAGTCCAGAGGCGCATCCCC 178399
QY 455 ThrGlyGluProValCysValYrThrAlaGluAspProAspLysGluAsnGlnLysIle 474
DB 178400 ACTGGGAGCCTGTGTGTCTACACTGCAGAGACCCCTGCACAGGAGATCAAAAGATC 178459

RESULT 11
AP002024
LOCUS      AP002024      54207 bp      DNA      linear      HTG 01-JUN-2000
DEFINITION Homo sapiens chromosome 16 clone 97G12 map 16q22, *** SEQUENCING IN
ACCESSION AP002024
VERSION    AP002024.1 GI:7798592
KEYWORDS   HTG; HTGS_PHASE2
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE      Tsai, S.F.
JOURNAL    Direct Submission
COMMENT    Submitted (09-MAY-2000) Shih-Feng Tsai, National Yang-Ming
          University, Institute of Genetics, 155 Li-Rong St. Section 2,
          Peitou, Taipei, Taiwan 11221, Republic of China
          (E-mail:lympet@aiym.edu.tw, URL:http://genome.ym.edu.tw/,
          Tel:886-2-28267043, Fax:886-2-28264930)
          Gaps between the contigs are represented as 100 N.
          * NOTE: This is a 'working draft', sequence. It currently
          * consists of 3 contigs. Gaps between the contigs
          * are represented as runs of N. The order of the pieces
          * is believed to be correct as given, however the sizes
          * of the gaps between them are based on estimates that have
          * provided by the submitter.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
          * 1 15198: contig of 15198 bp in length
          * 15199 15298: gap of 100 bp
          * 15299 25043: contig of 9745 bp in length
          * 25044 25143: gap of 100 bp
          * 25144 54207: contig of 29064 bp in length.
FEATURES   Location/Qualifiers
source     1..54207
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="16"
            /map="16q22"
            /clone="97G12"

ORIGIN
Alignment Scores:
Pred. No.:      3,09e-64      Length:      54207
Score:          72.00        Matches:      72
Percent Similarity: 100.00%    Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:    8.69%        Indels:      0

DB: 2 Gaps: 0
US-09-916-849A-1 (1-829) x AP002024 (1-54207)

QY 524 GlySerProThrThrThrGlyThrGlyThrLeuLeuThrLeuLeuAspValAsnAsp 543
DB 292 GGAAGCCCTCCACCACTGGCAGCGGAACCCCTTCTGTACACTGATTGAIGTCAACGAC 351
QY 544 HisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerProValArgHis 563
DB 352 CATGGCCCACTCCCTGAGCCCGCTCAGATCACCATCTGCAACCAAGCCCTGTGGCGCAC 411
QY 564 ValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGlnAlaGlnLeu 583
DB 412 GTGCTGAACATCAGGCAAGGACCTGTCTCCCCACACCTCCCTTTCAGGCCCGACGTC 471
QY 584 ThrAspAspSerAspIleTyrThrAlaGluVal 595
DB 472 ACAGTACTCAGACATCTACTGACGGCAGAGGTC 507

RESULT 12
BTCDHP
LOCUS      BTCDHP      2106 bp      mRNA      linear      MAM 20-JUL-1995
DEFINITION Bovine mRNA for P-cadherin.
ACCESSION X53614
VERSION    X33614.1 GI:166
KEYWORDS   cadherin; calcium binding protein; cell adhesion molecule;
          transmembrane protein.
SOURCE     Bos taurus (cow)
ORGANISM   Bos taurus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
          Bovidae; Bovinae; Bos.
REFERENCE  1 (bases 1 to 2106)
          Liaw, C.W., Cannon, C., Power, M.D., Kiboneka, P.K. and Rubin, L.L.
          Identification and cloning of two species of cadherins in bovine
          endothelial cells
          EXBO J. 9 (9), 2701-2708 (1990)
          90360979
          PUBMED 2390989
          POWER, M.D.
          Direct Submission
          Submitted (26-JUN-1990) Power M.D., Athena Neurosciences, 800 - F
          Gateway Blvd., South San Francisco, CA 94080, USA
          Data kindly reviewed (21-NOV-1990) by Power M.
FEATURES   Location/Qualifiers
source     1..2106
            /organism="Bos taurus"
            /mol_type="mRNA"
            /db_xref="taxon:9913"
            /cell_type="adrenal microvascular endothelial cell"
            /clone_lib="lambda gt10"
            /cdoon_start=3
            /product="P-cadherin (AA 232-723)"
            /protein_id="CAAJ37676.1"
            /db_xref="GI:833777"
            /db_xref="GOA:P19535"
            /db_xref="SWISS-PROT:P19535"
            /translation="ENTVSREVRLTVTDLDPNSPAWRATYRIVGGDNGDHEITITD
          PESNQGLTTQKGLDFAKTHLYEVINEVFVKLPSTATVVLVNEPVPVF
          VPSKVEIQEGISTGEPICATYARDPKGSKISYHILRDPAGWLAMDPSGQVTA
          GVDREDEQFVRNNIYEVMLATDGSPTTGTGTLTLLTMDNDHGFVPEQITTC
          NQSPVQVLNTDKLSPTAPFOALTSDVYVTAENVKGDVALSLKFLKQGE
          YDHLSTSDHGNKQLTVIRATVCTDCHGNMTCDPMTWGFLLPLILGAALLLLLV
          LFLVRKKRIKEPLLIPEDDTRNRYVEEGGGEEDODYDITOLHRLGARPEVVL
          RNDVAPSFITFMYRPPANPDEIGNFIENKANTDTPAPYDLSLVDFEGSGSD
          AASLSLTSSDQDQDYNILNEWGSRFKLADYVGGQDD"
            2088..2093
            /note="polyA signal"
            2106
            /note="polyA site"

```



```

/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:834535"
/tissue_type="Embryo, day 9 mouse (C57BL/6 background)
otocysts"
/clone_lib="NIH MGC_130"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6.1"
1..2993
/gene="Cdh3"
/notes="synonym: P-cadherin"
/db_xref="LocusID:12560"
/db_xref="MGI:88356"
1..2264
/gene="Cdh3"
/codon_start=3
/product="Cdh3 protein"
/protein_id="AAH52189.1"
/db_xref="GI:30704941"
/db_xref="LocusID:12560"
/db_xref="MGI:88356"
/translation="DIIMTRGTVOGGKDMHSPPTILRRKKRWMPPIFVPENGK
GFPPORLNQKSKDRGKIFYSITGPGADSPPEGVFTIEKESGWLILHMLDLREKIV
KYLELGHAVSENGASVEEPMNISIIVTDQNKPKFTQDITFRGSVLEGWPGTSMQV
TATDEDDAVNTYNGVAVSIHQEPKPHDLMTIHKSTGTSIVSSGLDREKPEYR
LTVDMDGEGSTTAAVAQIILDANDNAPEFEQKYEAWPENEVEHVEYRLTVD
LDVPSPAWRATYHIVGGDGDHFTITTHPTNQGLVTKKGLDPEAQDQTLXVEVT
NEAPAVKLPTATATVHVVDVNEAPVFPVPSKVIEAQEGISIGELVCIVTAQDPDK
EDQKSYLISRPANLAVDPSGOITRAGILDREDOFVKNVYEVWVLATDSGNPP
TTGTGLTLTLDINDHGPPEPRQIICNQSPPVQVNIITDKLSNPSFPQALTH
DSDIYMAEVSEKGTIVLSLFLKQDITDHLSDHGRNREQLMIRATVCDCHGQ
VENDCPWKGFIPLIGAVLALRLLALLLVKRVKPEKPELLPEDTRDNVYF
YEEGGEGEDQDITQLHGLRLEPVLRLNDVVFTIPTMYRPRANDEIGNFI
IENLKAANTDPTAPPYDSLLVFDYEGSGDAASLSLTSASDQDQDYNVNLWNGSRP
KILADWYGGGEDD"
105..389
/gene="Cdh3"
/notes="cadherin; Region: Cadherin domain"
/db_xref="CDD:pfam00028"
480..749
/gene="Cdh3"
/notes="CA; Region: Cadherin repeats. Cadherins are
glycoproteins involved in Ca2+-mediated cell-cell
adhesion. Cadherin domains occur as repeats in the
extracellular regions which are thought to mediate
cell-cell contact when bound to calcium"
/db_xref="CDD:smart00112"
768..1064
/gene="Cdh3"
/notes="cadherin; Region: Cadherin domain"
/db_xref="CDD:pfam00028"
1158..1409
/gene="Cdh3"
/notes="CA; Region: Cadherin repeats. Cadherins are
glycoproteins involved in Ca2+-mediated cell-cell
adhesion. Cadherin domains occur as repeats in the
extracellular regions which are thought to mediate
cell-cell contact when bound to calcium"
/db_xref="CDD:smart00112"
1422..1685
/gene="Cdh3"
/notes="cadherin; Region: Cadherin domain"
/db_xref="CDD:pfam00028"
1845..2252
/gene="Cdh3"
/notes="Cadherin C term; Region: Cadherin cytoplasmic
region. Cadherins are vital in cell-cell adhesion during
tissue differentiation. Cadherins are linked to the
cytoskeleton by catenins. Catenins bind to the cytoplasmic
tail of the cadherin. Cadherins cluster to form foci of
homophilic binding units. A key determinant to the
strength of the binding that it is mediated by cadherins
is the juxtamembrane region of the cadherin. This region

```

```

induces clustering and also binds to the protein p120ctn"
/db_xref="CDD:pfam01049"

ORIGIN
Alignment Scores:
Pred. No.: 4,28e-43 Length: 2993
Score: 51.00 Matches: 51
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 6.15% Indels: 0
DB: 10 Gaps: 0

US-09-916-849A-1 (1-829) x BC052189 (1-2993)

Qy 684 LysGluProLeuLeuLeuProGluAspThrArgAspAsnValPheTyrTyrGlyGlu 703
Db 1824 AAGAGCCCTTCTGCTCCAGAGATGACACCGGACACATGCTCTTATTATGAGAA 1883
Qy 704 GluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArgGlyLeuGlu 723
Db 1884 GAGGCTGTGTGTAAGAGGACCGAGGACTATGACATCACCACTCCACCGGAGCTGGAG 1943
Qy 724 AlaArgProGluValValLeuArgAspVal 734
Db 1944 GCCAGGCTGAGGTGGTCTCCGAACGAGTA 1976

RESULT 15
LOCUS MWCADHP 3187 bp mRNA linear ROD 30-MAR-1995
DEFINITION Mouse mRNA for P-cadherin.
ACCESSION X06340.1 GI:50267
VERSION X06340.1
KEYWORDS cadherin; cell adhesion molecule; glycoprotein; transmembrane
protein.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 3187)
AUTHORS Nose,A., Nagafuchi,A. and Takeichi,M.
TITLE Isolation of placental cadherin cDNA: identification of a novel
gene family of cell-cell adhesion molecules
JOURNAL EXBO J. 6 (12), 3655-3661 (1987)
MEDLINE 88111554
PUBMED 3428270
FEATURES
source
1..3187
Location/Qualifiers
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="P3"
/cell_line="PSA5-E"
46..2514
/notes="unnamed protein product; pre P-cadherin (AA -27 to
795)"
/codon_start=1
/protein_id="CAA29646.1"
/db_xref="GI:50268"
/db_xref="GOA:P10287"
/db_xref="SWISS-PROT:P10287"
/translation="MELLSGFHAPLLLLQLQVLRWSVSPYPAGFTIGRAGVTLVEG
TDLEPSOVLKVALAGQMHADNGDIIMLTRGTVOGGKDMHSPPTILRRKKRW
MPPIFVPENGKGPFPQNLQKSKDRGKIFYSITGPGADSPPEGVFTIEKESGWL
LHMLDREKIVKYLELGHAVSENGASVEEPMNISIIVTDQNKPKFTQDITFRGSV
LWPGTSMQVQVATDEDDAVNTYNGVAVSIHQEPKPHDLMTIHKSTGTSIVSS
GLDREKPEYRLTVDMDGEGSTTAAVAQIILDANDNAPEFEQKYEAWPENEVEH
GHEVRLTVDLDVNEAPFAVKLTATATVHVVDVNEAPVFPVPSKVIEAQEGISIG
ELVCIVTAQDPDKQKISYTSRDPANLAVDPSGOITRAGILDREDOFVKNVYEV
MVLATDSGNPTTGTGLTLLTLDINDHGPPEPRQIICNQSPPVQVNIITDKLSN
NSSPQALTHDSDIYMAEVSEKGTIVLSLFLKQDITDHLSDHGRNREQLMIRAT
IRATVCDCHGVENDCPWKGFIPLIGAVLALRLLALLLVKRVKPEKPELLPEDTR
PEDTRDNVYFEEGGEGEDQDITQLHGLRLEPVLRLNDVVFTIPTMYRPRANDEI

```

CDS

PANPDEIGNFIENLKPANTDPTAPPYDSLWVFDYEGSGDAASLSLTTSASDQDQD
 YNVLNWSGRFKKLADMYGGEDD"

sig_peptide 46. 126 /note="signal peptide (AA -27 to -1)"
 mat_peptide 127. .2511 /product="P-cadherin (AA 1 - 795)"
 misc_feature 619. .621 /note="pot. N-linked glycosylation site"
 misc_feature 1690. .1692 /note="pot. N-linked glycosylation site"
 misc_feature 1717. .1719 /note="pot. N-linked glycosylation site"
 misc_feature 1744. .1746 /note="pot. N-linked glycosylation site"
 misc_feature 3164. .3170 /note="pot. N-linked glycosylation site"
 misc_feature /note="polyA signal"

ORIGIN

Alignment Scores:
 Pred. No.: 4.54e-43 Length: 3187
 Score: 51.00 Matches: 51
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 6.15% Indels: 0
 DB: 10 Gaps: 0

US-09-916-849A-1 (1-829) x MNCADHP (1-3187)

QY 684 LysGluProLeuLeuProGluAspThrArgAspAsnValPheTyrTyrGlyGlu 703
 Db 2074 AAAGAGCCCTTCTGCTCCAGAGATGACACCGGAGACAATGCTTCTATTATGAGAA 2133
 QY 704 GluGlyGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisArgGlyLeuGlu 723
 Db 2134 GAGGGTGGTGGTGAAGAGGACCAGGACTATGACATCACCACCACTCCACCGGGGACTGGAG 2193
 QY 724 AlaArgProGluValValLeuArgAsnAspVal 734
 Db 2194 GCCAGCCCTGAGTGGTGTCTCCGAAACAGATGA 2226

Search completed: September 23, 2004, 06:21:07
 Job time : 8604 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 23, 2004, 01:24:55 ; Search time 768 Seconds
(without alignments)
4585.620 Million cell updates/sec

Title: US-09-916-849A-1

Perfect score: 829

Sequence: 1 MGLPRGPIASLLLLQVCWLQ.....NEWGSRFKLADMYGGGDD 829

Scoring table: OLIGO

Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Word size: 1

Total number of hits satisfying chosen parameters: 6745146

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n.model -DEV=xlh
-Q/cn2_1/USPTO_spooh_h/US0916849/runat_21092004_145834_7311/app_query.fasta_1.967
-DB=N Geneseq_29Jan04 -QFMT=fastap -SUFFIX=oligo.rng -MINMATCH=0.1 -DOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0916849 @CGN 1 1 513 @runat_21092004_145834_7311 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT_DSPELOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N Geneseq_29Jan04:*

1: Geneseqm1980s:.*
2: Geneseqm1990s:.*
3: Geneseqm2000s:.*
4: Geneseqm2001as:.*
5: Geneseqm2001bs:.*
6: Geneseqm2002as:.*
7: Geneseqm2003as:.*
8: Geneseqm2003bs:.*
9: Geneseqm2003cs:.*
10: Geneseqm2004s:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	829	100.0	3171	6 ABK92214	Abk92214 Prostate
2	829	100.0	3171	6 ABN89393	Abn89393 Human P-c
3	829	100.0	3171	6 ABTC7755	Abtc7755 Breast ca
4	829	100.0	3171	6 ABK52038	Abk52038 DNA encod
5	829	100.0	3171	6 ABN97264	Abn97264 Gene #376
6	829	100.0	3171	7 ACAS6750	Acas6750 Human sig
7	829	100.0	3171	7 ABZ24736	Abz24736 Human p-c
8	829	100.0	3171	9 ADD14795	Add14795 Human src

9	829	100.0	3777	8 ACH03801	Ach03801 Human cdn
10	728	87.8	3205	6 ABQ81547	Abq81547 Gene up-r
11	728	87.8	3205	7 ACC72821	Acc72821 Human can
12	728	87.8	3205	7 ABX76155	Abx76155 Lung canc
13	728	87.8	3205	7 ABX76399	Abx76399 Lung canc
14	715	86.2	3219	6 AAD27637	Aad27637 Human col
15	102	12.3	306	5 AAH81822	Aah81822 Human dif
16	61	7.4	2156	2 AAQ11562	Aaq11562 Partial e
17	54	6.5	255	5 AAH81872	Aah81872 Rat diffe
18	52	6.3	160	7 ABZ24737	Abz24737 Oligonuc
19	21	2.5	522	6 ABK53882	Abk53882 Human bea
20	19	2.3	60	6 ABN40851	Abn40851 Human sgl
21	16	1.9	2310	7 ABZ09877	Abz09877 Human 5'
22	16	1.9	5310	9 ADE84014	Ade84014 5' regula
23	15	1.8	2768	6 ABN89352	Abn89352 Mouse E-c
24	15	1.8	2768	6 ABN89353	Abn89353 Mouse E-c
25	15	1.8	4333	2 AAQ11563	Aaq11563 Encodes E
26	15	1.8	4396	7 ABT42198	Abt42198 Toxicity
27	15	1.8	4396	2 ADB53244	Adb53244 Primary r
28	15	1.8	4778	2 AAX56022	Aax56022 Wild-type
29	15	1.8	4778	4 AAF86687	Aaf86687 Human E-c
30	15	1.8	4828	6 ABS76385	Abs76385 cDNA enco
31	15	1.8	4828	6 ABV94337	Abv94337 Breast ca
32	14	1.7	2808	2 AAT05764	Aat05764 Human E-c
33	13	1.6	317	5 ABV19790	Abv19790 Human pro
34	13	1.6	484	5 ABV49555	Abv49555 Human pro
35	13	1.6	2808	6 ABN89392	Abn89392 Human E-c
36	13	1.6	2815	2 AAQ65487	Aaq65487 Sequence
37	13	1.6	3048	2 AAT85433	Aat85433 Human cad
38	13	1.6	3048	2 AAT61920	Aat61920 Full leng
39	13	1.6	3048	6 ABK83876	Abk83876 Human cdn
40	13	1.6	3048	9 ADD14706	Add14706 Human src
41	13	1.6	3310	9 ADB62391	Adb62391 Human cdn
42	12	1.4	403	8 ACHI7807	Achi7807 Human adu
43	12	1.4	436	6 ABT07156	Abt07156 Human ova
44	12	1.4	436	6 ABX73034	Abx73034 Human ova
45	12	1.4	467	7 ABK54052	Abk54052 Human hea

RESULT 1

ABK92214
ID ABK92214 standard; DNA; 3171 BP.

XX AC ABK92214;

XX DT 15-AUG-2002 (first entry)

XX DE Prostate cancer-associated DNA sequence #100.

XX KW Prostate cancer; prostate tumour tissue; human; mammal; cytostatic;

XX KM gene therapy; gene; ds.

XX OS Mammalia.

XX FN WO200203268-A2.

XX PD 18-APR-2002.

XX PF 12-OCT-2001; 2001WO-US032045.

XX PR 13-OCT-2000; 2000US-00687576.

XX PR 08-DEC-2000; 2000US-00733288.

XX PR 08-DEC-2000; 2000US-00733742.

XX PR 24-JAN-2001; 2001US-0263957P.

XX PR 16-MAR-2001; 2001US-0276791P.

XX PR 06-APR-2001; 2001US-0276888P.

XX PR 24-APR-2001; 2001US-0281922P.

XX PR 30-APR-2001; 2001US-0286214P.

XX PR 04-MAY-2001; 2001US-00847046.

XX XX 04-MAY-2001; 2001US-0288589P.

ALIGNMENTS

[illegible]DE Human P-cadherin encoding cDNA SEQ ID NO:49.
XX
XX Cadherin; alpha-catenin; cancer; beta-catenin binding domain; melanoma;
KW o-catenin; colon cancer; gene; ss.
XX Homo sapiens.
OS
XX
FH Key Location/Qualifiers
FT CDS 54..2543
FT /*tag= a
FT /product= "P-cadherin"
XX US2002045591-A1.
PN
PD 18-APR-2002.
XX
XX 17-JUL-2001; 2001US-00905983.
XX
PR 26-MAY-1998; 98IL-00124650.
PR 26-MAY-1999; 99US-00318633.
XX
PA (GEIG/) GEIGER B.
PA (BENZ/) BEN-ZE'EV A.
PA (SADO/) SADOT E.
XX
PI Geiger B, Ben-Ze'ev A, Sadot E;
XX
XX WPI; 2002-439105/53.
DR P-FSDB; ABB81476.
XX
PT New construct encoding soluble cytoplasmic portion of cadherin including
PT beta catenin binding domain useful in treating cancer associated with
PT high beta-catenin activity e.g. colon cancer and melanoma.
XX
PS Claim 3; Page 51-53; 102pp; English.
XX
CC The present invention describes a pharmaceutical composition for treating
CC cancer associated with abnormally high beta-catenin activity. The
CC pharmaceutical composition comprises a gene therapy vehicle harbouring a
CC polynucleotide that contains: (i) a nucleotide sequence encoding a
CC soluble cytoplasmic portion of a cadherin which lacks a transmembrane
CC portion and an extracellular portion of the cadherin, and includes a beta
CC -catenin binding domain; and (b) an upstream promoter for directing
CC expression of the soluble cytoplasmic portion of the cadherin in a
CC mammalian cell. Also described is a pharmaceutical composition for
CC treating cancer associated with abnormally high activity levels of beta-
CC catenin comprising a gene therapy vehicle harbouring a polynucleotide
CC that contains: (a) a nucleotide sequence encoding an o-catenin; and (b)
CC an upstream promoter for directing expression of the o-catenin in a
CC mammalian cell. The pharmaceutical compositions have cycostatic activity
CC and can be used in the suppression of beta-catenin-mediated
CC transactivation. They can be used for treating cancers associated with
CC abnormally high activity levels of beta-catenin such as colon cancers and
CC melanomas, by reducing these high activity levels of beta-catenin in
CC mammalian cells. The present sequence encodes human P-cadherin which is
CC used in the exemplification of the present invention
SQ Sequence 3171 BP; 740 A; 903 C; 864 G; 664 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 0 Length: 3171
Score: 829.00 Matches: 829
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-916-849A-1 (1-829) x ABN89393 (1-3171)

Cy 1 MetGlyLeuProArgGlyProLeuAlaSerIeuLeuleuLnuValCysTrpLeuCln 20

Db 54 ATGGGGTCCTCGTGAGACCTCTCGGTCTCTCTTCTCCAGGTTTCGTGCCTCGAG 113

21 CysAlaAlaSerGluProCysArgAlaValPheAraGluAlaGluValThrLeuGluAla 40
Db
114 TGGCGGCGCTCCAGCGCTGCGGCGGTCTTCAGGGAGGTGAAGTACCTTGGAGGCG 173
Qy
41 GlyGlyAlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
Db
174 GGAGGCGCGAGCAGGAGCGCGCGAGCGCTTGGGGAAGTATTATGGGCTGCCCTGGG 233
Qy
61 GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValAraCAsnGlyGluThr 80
Db
234 CAAGAGCCAGCTCTGTTAGCACTGATATATGATGACTTCACTGTCGGGAATGGCGAGACA 293
Qy
81 ValGlnGluArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
Db
294 GTCCAGAAAGAGGTCACTGAAGAAAGAAATCCATTGAAGATCTTCCCATCAAAAGT 353
Qy
101 IleLeuArgArgHisLysArgAspTrpValAlaProIleSerValProGluAsnGly 120
Db
354 ATCTTACGAGACACAGAGAGATTGGGTGGTTGCTCCAATATCTGTCCCTGAATATGGC 413
Qy
121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
Db
414 AAGGGTCCCTTCCCCAGAGACTGAATCAGCTCAAGTCTAATAAAGATAGAGACCAAG 473
Qy
141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProGluGlyValPheAlaVal 160
Db
474 ATTTTCTACAGCATCACGGGGCGGGGCGAGACGCCCTCCCTGAGGGTGTCTTGCTGTA 533
Qy
161 GluLysGluThrGlyTrpLeuLeuLeuLysProLeuAspArgGluGluIleAlaLys 180
Db
534 GAGAGGAGACAGGCTGGTTGTTGTAATAGCCACTGGACCGGGAGAGATTGCCAAG 593
Qy
181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
Db
594 TATGAGCTCTTTGGCCACGCTGTGTGAGAGATGGTGGCTCAGTGGAGGAGCCCATGAAC 653
Qy
201 IleSerIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
Db
654 ATCTCCTCATCGTACCACAGCAATGACCAAGCCCAAGTTCATCCAGGACACCTTC 713
Qy
221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
Db
714 CGAGGGAGTGTCTTGAAGGAGTCTCTACAGGTACTTCTGTGATGCAGGTGACAGCCACA 773
Qy
241 AspGluAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln 260
Db
774 GATGAGGATGATGCCATCTACACCTACAAATGGGGTGGTGTCTTACTCCATCCATAGCCAA 833
Qy
261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
Db
834 GAACCAAGAGACCCACACACCTCATGTTCACAATTCACGGAGCACAGGCACCATCAGC 893
Qy
281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
Db
894 GTCATCTCCAGTGGCGTGACCGGAAAGTCCCTGAGTACACACTGACCATCCAGGCC 953
Qy
301 ThrAspValAspGlyAspGlySerThrThrAlaValAlaValAlaValGluIleLeuAsp 320
Db
954 ACAGACATGGATGGGACCGGCTCCACCCACCGAGCTGGCGAGTAGTGGAGATCCTTGAT 1013
Qy
321 AlaAsnAspAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
Db
1014 GCCAATGCAATGCTCCCATGTTTACCCCAAGAGTACGAGGCCCATGTGCTGAGAAAT 1073
Qy
341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
Db
1074 GAGATGGGCGCATGAGGTGAGAGGCTGACGGTCACTGATCTGGACGCCCCCACTCACCA 1133
Qy
361 AlaTrpArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
Db
1134 GCGTGGCGTGGCCACTTATCTTATGCGCGGTGACGCGGGACCAATTTTACCATCACC 1193
Qy
381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLysPheIleLeuGlu 400

1194 ACCACCTTGAGAGCAACAGGGCACTCTGCAACACCGAGGAAGGTTTGGATTTTGAGGCC 1253
Qy
401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db
1254 AAAACCCAGCACACCTCTGACGTTGAAGTGACCAACAGAGCCCTTTTGTGCTGAAGCTC 1313
Qy
421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
Db
1314 CCAACCTCCACAGCCACCATAGTGTCCACGTGGAGGATGTGAATGAGGCACCTGTGTTT 1373
Qy
441 ValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
Db
1374 GTCCCACTCTCCAAAGTGTGTGAGTCCAGGAGGGCATCCCCACTGGGAGGCTGTGTGT 1433
Qy
461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db
1434 GTCTACACTGCAGAGACCTCTGACAGGAGAAATCAAAAGATCAGCTACCGCATCCTGAGA 1493
Qy
481 AspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db
1494 GACCCAGCAGGGTGGCTAGCCATGGACCCAGACAGTGGGCGAGTCCACAGCTGTGGGCACC 1553
Qy
501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db
1554 CTGACCGTGAAGATGAGCAGTTTGTGAGGAACAACATCTATGAAGTCAATGTCTTGGCC 1613
Qy
521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAsp 540
Db
1614 ATGGACAATGTAAGAGCTCTCCACCACTGGCAAGGAAACCTCTCTGTAACACTGATGAT 1673
Qy
541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db
1674 GTCAACGACCATGGCCAGTCCCTGAGCCCCGTGAGTCACTCAGTCACTGCAACCAAGCCCT 1733
Qy
561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
Db
1734 GTGGGCCACGTGTGAACATCAAGCAAGGACCTGTCTCCCAACACCTCCCTTCCAG 1793
Qy
581 AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGlyAsp 600
Db
1794 GCCCAGCTCACAGATGACTCAGACATCTACTGGACGGCAGAGTCAACGAGGAAGTGAAC 1853
Qy
601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
Db
1854 ACAGTGGTCTTGTCTCTGAGAGAGTCTCTGAGCAGGATACATATGACGTGCACCTTCT 1913
Qy
621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
Db
1914 CTGTCTGACCATGGCAACAAAGAGCAGCTGACGGTGTATCAGGGCCACTGTGTGCGACTGC 1973
Qy
641 HisGlyHisValGluThrCysProGlyProTyrLysGlyGlyPheIleLeuProValLeu 660
Db
1974 CATGGCCATGTGAAACCTGCCCTGGACCTCGAAGAGAGGTTTCACTCTCTCTGTGCTG 2033
Qy
661 GlyAlaValLeuAlaLeuLeuPheLeuLeuLeuValLeuLeuLeuValArgLysLys 680
Db
2034 GGGGCTGTCTGTGCTCTGTGTCTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2093
Qy
681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
Db
2094 CGAAAGATCAAGGAGCCCTCTCTACTCCAGAGATGACACCCGTGACACCTCTCTTAC 2153
Qy
701 TyrGlyGluGlyGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
Db
2154 TATGGCGAGAGGGGGTGGCGAAGAGGACAGGACTATGACATCACCCAGCTCCACCGA 2213
Qy
721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleLeuPro 740
Db
2214 GGTCTGAGGCCAGCGGAGGTGTCTCCCAATGACGTGGCAGCAACCACTATCCCG 2273
Qy
741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleLeuGlu 760

1674 GTCAAGACCATGGCCAGTCCCTGAGCCCGTCAGATCACCATCTGCAACCAAGCCCT 1733
 QY ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
 1734 GTCCGCCAGCTGCTGAACATCAGGACAGGACCTGTCTCCCCACACCTCCCTTTCCAG 1793
 QY AlaGlnLeuThrAspAspSerAspIleTyrThrAlaGluValAsnGluGluGlyAsp 600
 1794 GCCAGCTCACAGATGACTCACATCTACTGAGCGCAGAGGTCAACGAGGAGGTGAC 1853
 QY ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
 1854 ACAGTGGTCTTCTGCTGAAGAAGTCTCTGAAGCAGGATACATATGAGCTGCACCTTTCT 1913
 QY LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
 1914 CTGTCTGACCATGGCACAAGAGCAGCTGACGGTGATCAGGGCCACTGTGTGGACTGC 1973
 QY HisGlyHisValGluThrCysProGlyProTyrProTyrLysGlyGlyPheIleLeuProValLeu 660
 1974 CATGCCATGTCAAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2033
 QY GlyAlaValLeuAlaLeuLeuPheLeuLeuLeuValLeuLeuValArgLysLys 680
 2034 GGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2093
 QY ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
 2094 CGGAAGATCAAGGAGCCCTCTACTCCAGAGATGACACCGGTGACACAGTCTTCTAC 2153
 QY TyrGlyGluGluGlyGlyGlyGluGluAspGlnAspTyrAspIleThrGluLeuHisArg 720
 2154 TATGCGAAGAGGGGGTGGGAGAGGAGGACCGAGTATGACATCACCAGCTCCACCGA 2213
 QY GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
 2214 GGTCTGGAGCCAGCGCGAGTGTCTTCCGATGACGTGGCACCACCATCATCCCG 2273
 QY ThrProMetTyrArgProArgProAlaSerProAspGluIleGlyAsnPheIleGlu 760
 2274 ACACCATGTACCGTCTAGGCGCAGCAGCAGCAGATGAATCGCACTTTATATTTGAG 2333
 QY AsnLeuLysAlaAlaAsnThrAspProThrAlaProTyrAspThrLeuLeuValPhe 780
 2334 AACCTGAAGCGGCTTAACACACCCACAGCCCGCCCTAGCAGCCTCTTGTGTTC 2393
 QY AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
 2394 GACTATGAGGCGAGCGGCTCCGACGCGGCTCCCTGAGCTCCCTCACCTCTCCGCTCC 2453
 QY AspGlnAspGlnAspTyrAspTyrLeuGlnGluTyrGlySerArgPheLysLeuAla 820
 2454 GACCAAGACCAAGATTACCATTTATCTGAACAGTGGGCGGCGCTTCAAGAGCTGSCA 2513
 QY AspMetTyrGlyGlyGluAspAsp 829
 2514 GACATGTACGGTGGCGGAGGAGCAGC 2540

RESULT 5
 ABN97264
 ID ABN97264 standard; DNA; 3171 BP.
 XX
 AC ABN97264;
 XX
 DT 13-AUG-2002 (first entry)
 XX
 DE Gene #3762 used to diagnose liver cancer.
 XX
 KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
 KW metastatic liver tumor; cytostatic; expression profile; disease state;
 KW disease progression; drug toxicity; drug efficacy; drug metabolism.
 XX
 OS Homo sapiens.

XX WO200229103-A2.
 XX
 PD 11-APR-2002.
 XX
 PF 02-OCT-2001; 2001WO-09030589.
 XX
 PR 02-OCT-2000; 2000US-0237054P.
 XX
 PA (GENE-) GENE LOGIC INC.
 XX
 FI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
 XX WPI; 2002-426119/45.
 DR
 XX
 XX
 PT Diagnosing and detecting the progression of liver cancer, hepatocellular
 PT carcinoma or metastatic liver tumor in a patient, involves detecting the
 PT level of expression of two or more genes in a liver tissue sample.
 XX
 PS Claim 1; SEQ ID NO 3762; 298pp; English.
 XX
 CC The invention relates to a novel method for diagnosing and detecting the
 CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
 CC tumor in a patient, and differentiating metastatic liver cancer from
 CC hepatocellular carcinoma in a patient, involving detecting the level of
 CC expression of two or more genes represented in ABN93503-ABN97455 in a
 CC tissue sample. The method of the invention has hepatotropic, and
 CC cytotatic activity. The method is useful for diagnosing and detecting
 CC the progression of liver cancer, hepatocellular carcinoma and metastatic
 CC liver carcinoma in a patient. The method is useful for identifying
 CC expression profiles which serve as useful diagnostic markers as well as
 CC markers that can be used to monitor disease states, disease progression,
 CC drug toxicity, drug efficacy and drug metabolism. Note: The sequence data
 CC for this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 3171 BP; 740 A; 903 C; 864 G; 664 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 0 Length: 3171
 Score: 829.00 Matches: 829
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DE: Gaps: 0

US-09-916-849A-1 (1-829) x ABN97264 (1-3171)
 QY 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuGlnValCysTrpLeuGln 20
 Db 54 ATGGGCTCCCTCGTGGACCTCTCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 113
 QY 21 CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
 Db 114 TGGCGCGCTCCGAGCGCTGCCGCGCGCTCTTCAGGAGCGCTGAGTACCTTGGAGCG 173
 QY 41 GlyGlyValaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
 Db 174 GGAGGCGCGGAGCAGAGCGCCGCGCGCTGGGGAAGTATTTCATGGCTGCCCTGGG 233
 QY 61 GlnGluProAlaLeuPheSerThrAspAsnAspAspPheThrValArgAsnGlyGluThr 80
 Db 234 CAAGAGCCAGCTCTGTTTAGCACTGATGATGATGATGATGATGATGATGATGATGATG 293
 QY 81 ValGlnGluArgArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
 Db 294 GTCCAGGAAGAAGGTCACTGAGGAGGAAGGATCCATTGAAGATCTTCCATCCAAAGCT 353
 QY 101 IleLeuArgArgHisLysArgAspTrpValValAlaProIleSerValProGluAsnGly 120
 Db 354 ATCTTACGAAGACACAAAGAGATTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 413

121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProGluGlyValPheAlaVal 160
142 ATTTCTACAGCATCACGGGGCGGGGCGAGACAGCCCCCTGAGGGTGTCTTGCTGTA 533
161 GluLysGluThrGlyTyrLeuLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys 180
162 GAGAGAGGACAGGCTGGTGTGTGTGTAATAAGCCACTGGACCGGAGAGAGATTGCCAAG 593
181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
182 TATGAGCTCTTTGGCCACGCTGTGTACAGAGATGGTGCTCAGTGGAGGACCCCATGAAC 653
201 IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGluAspThrPhe 220
202 ATCTCATCATCTGACCGACCGAGATGACCAAGCCCAAGTTTACCAGGACACCTTC 713
221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
222 CGAGGGAGTGTCTTAGAGGAGTCTTACAGGTACTTCTGTGATGCAGGTGACGCCACA 773
241 AspGluAspAspAlaIleTyrThrAsnGlyValValAlaTyrSerIleHisSerGln 260
242 GATGAGGATGATGCCATCTACACCTACCAATGGGGTGGTGTCTTACTCCATCCATAGCCAA 833
261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
262 GAACCAAGGACCCACACGACCTCATGTTCACATTCACCGGAGCACAGCCACCATCAGC 893
281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
282 GTTCATCTCCAGTGGCTGGACCGGAAAAGTCCCTGAGTACACACTGACCATCCAGGCC 953
301 ThrAspMetAspGlyAspGlySerThrThrAlaValAlaValAlaValGluIleLeuAsp 320
302 ACAGACATGATGGGACGGCTCCACCCACCGGACGAGTGGCAGTAGTGGAGATCCTTGAT 1013
321 AlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
1014 GCCAATGACAAATGCTCCCATGTTTACCCCAAGAAAGTACGAGGCCCATGTGCTGAGAA 1073
341 AlaValGlyHisGluValClnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
1074 GCAGTGGGCGCATGAGGTGAGAGGTGACGGTCACTGATCTGGACGCCCCCACTCACA 1133
361 AlaTyrArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
1134 CGTGGGCGTCCACCTACCTTATCATGGCGGTGACGACGCGGACCATTTTACCATCACC 1193
381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
1194 ACCCACTTGAGAGCAACCGAGGCCATCTGTGACCAACAGGAGGGTGTGGATTTTGGAGCC 1253
401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
1254 AAAAACCCAGCACACCTGTACCTGTGAGTGCACCAACGAGGCCCTTTTGTGCTGAGCTC 1313
421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
1314 CCAACCTCCACAGCCACCATAGTGTGTCACGTGGAGGATGTGAATGAGGCCACTGTGTT 1373
441 ValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
1374 GTCCACCTCCAAAGTCGTGAGGTCCAGGAGGAGTCCCCACCTGGGAGCCTGTGTGT 1433
461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
1434 GTCTACACTGCAGAGACCTTCACAGGAGAGATCAAAAGATCAGCTACCGCATCTGAGA 1493
481 AspProAlaGlyTyrTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500

1494 GACCCAGCAGGGTGGGTAGCCATGGACCCACAGACAGTGGGCGAGTCAAGCTGTGGGCACC 1553
501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
1554 CTCACCGTGGAGATGAGCAGTTTGTGAGAAACAATCTATGAGTCAATGGTCTTGGCC 1613
521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAsp 540
1614 ATGGACAATGGAAGCCCTCCACCACTGGCAGCGGAACCCCTCTGCTAACACTGATTGAT 1673
541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
1674 GTCAACGACCATGGCCAGTCCCTGAGCCCGTCAGATCACCATCTGCACCAACCAAGCCCT 1733
561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
1734 GTGGCCCGCTGTGTAACATCAGGACAGGACCTGTCTCCCCACACCTCCCTCTCCAG 1793
581 AlaGlnLeuThrAspAspSerAspIleTyrTyrThrAlaGluValAsnGluGluAsp 600
1794 GCCCAGCTCAGATGACTCAGACATCTACTGACGCGCAGAGGTCAACGAGGAGGTGAC 1853
601 ThrValValLeuSerLeuLysLysPheLeuLysGluAspThrTyrAspValHisLeuSer 620
1854 ACAGTGTCTTGTCCCTGAAGAAGTTCTCGAAGCAGGATACATATGACGTGCACTTCT 1913
621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
1914 CTGTCTGACCATGGCAACAAAGACGAGCTCAGCGTGATCAGGGCCACTGTGTGCGACTGC 1973
641 HisGlyHisValGluThrCysProGlyProTyrGlyGlyGlyPheIleLeuProValLeu 660
1974 CATGGCCATCTGCAAACTCCCTCGACCTGGACCTGAAAGGAGTTTCACTCTCTGCTG 2033
661 GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuLeuValArgLysLys 680
2034 GGGCTGTCTGTGCTGTCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2093
681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
2094 CGGAAGATCAAGAGGCCCTCTACTCCCAAGAGATGACACCGGTGACAAACGTCTCTCTAC 2153
701 TyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
2154 TATGGCGAAGAGGGGGTGGCGAAGAGGACCAAGCTATGACATCACCCAGCTCCACCGA 2213
721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
2214 GGTCTGAGGCCAGCGCGAGGTGGTTCCTCGCAATGACGCTGGCACCAACCATCATCCCG 2273
741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
2274 ACACCATGTACCGTCTTAGCCAGCCCAACCCAGATGAATCGCAACTTTATTAATGAG 2333
761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProTyrAspThrLeuLeuValPhe 780
2334 AACCTGAAGCGCGTAAACACAGACCCCCACAGCCCGCCCTACGACACCTCTTGGTGTTC 2393
781 AspTyrGluGlySerGlySerAspAlaAlaSerIleSerSerLeuThrSerSerAlaSer 800
2394 GACTATGAGGGCAGCGGCTCCGACCGCGCTCCCTGAGCTCCCTCAGCTCCCTCCGCGCTCC 2453
801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTyrGlySerArgPheLysLysLeuAla 820
2454 GACCAAGACCAAGATTACGATTATCTGAACAGTGGGCGAGCGCTTCAAGAGCTGGCA 2513
821 AspMetTyrGlyGlyGlyGluAspAsp 829
2514 GACATGTACGTGGCGGGAGGACGAC 2540
RESULT 6
ACA56750

Db 1194 ACCACCTGAGACACAGGCGATCTGACACAGAGGGTTTGGATTGAGGCC 1253
Qy LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db 1254 AAAAAACACGACACACCTGTACGTTGAGTGACCAACGAGGCCCTTTGTGCTGAGGCTC 1313
Qy 421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
Db 1314 CCAACCTCCACAGCCACCATAGTGTCTCCAGTGGAGGATGGAATGAGGCACCTGTGT 1373
Qy 441 ValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
Db 1374 GTCCCAACCTCCAAAGTCGTTGAGGTCCAGGAGGCATCCCACTGGGGACCTGTGT 1433
Qy 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db 1434 GTCTACATGTGACAGACCCCTGACAGGAGATCAAAAGATCAGCTACCGCATCTGAGA 1493
Qy 481 AspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db 1494 GACCCAGCAGGGTGTCTAGCCATGACCCACAGACAGTGGCGAGGTACAGCTGTGGCACC 1553
Qy 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db 1554 CTCGACCGTGGAGATGAGCAGTTTGTGAGGACAAACATCTATGAAGTCATGCTCTGGCC 1613
Qy 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuThrLeuIleAsp 540
Db 1614 ATGGCAATGGAGCCCTCCACCACTGGCAGGGAACCTTCTGCTAACACTGATTGAT 1673
Qy 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 1674 GTCAACGACCATGGCCAGTCCCTGAGCCCGCTGAGATCACCATCTGCAACCAAGCCCT 1733
Qy 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
Db 1734 GTGCGCCACGTCGCTGAACATCAGGACAGGACCTGTCTCCCAACACCTCCCTTCCAG 1793
Qy 581 AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGluGlyAsp 600
Db 1794 GCCCAGCTCAGAGTACTCAGACATCTCTGGACGGCAGAGGTCAACGAGAGGTGAC 1853
Qy 601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
Db 1854 ACAGTGGTCTTGTCCCTGAAGAAGTCTCTGAACGAGGATACATATGACGTGACCTTCT 1913
Qy 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
Db 1914 CTGCTGACCATGGCCAAACAGAGCAGCTGACGGTGATCAGGCCACTGTGTGACATGC 1973
Qy 641 HisGlyHisValGluThrCysProGlyProTyrLysGlyPheIleLeuProValLeu 660
Db 1974 CATGGCCATGTGAAACCTGGACCTGGACCTGGAAAGGAGGTTTCATCTCCCTGTGCTG 2033
Qy 661 GlyAlaValLeuAlaLeuPheLeuLeuLeuValLeuLeuLeuValValArgLysLys 680
Db 2034 GGGGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2093
Qy 681 ArgLysIleLysGluProLeuLeuLeuProGluAspThrArgAspAsnValPheTyr 700
Db 2094 CGGAGATCAAGAGGCCCTCTACTCCAGAGATGACACCCGTGACAAAGCTCTCTAC 2153
Qy 701 TyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
Db 2154 TATGGCGAAGAGGGGGTGGCGAGAGGACCATGATGATCATCACCGAGTCCACCGA 2213
Qy 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleLeuPro 740
Db 2214 GGTCTGGAGGCCAGCGGAGGTGTTTCTCCCAATGACGTGGCCACCAACCATCATCCCG 2273
Qy 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleGlu 760

Db 2274 ACACCCATGTACCTGCTAGGCCAGCCCAACCCAGATGAATCGCACTTATTAATTGAG 2333
Qy 761 AsnLeuLysAlaAlaAsnThrAspPProThrAlaProProTyrAspThrLeuLeuValPhe 780
Db 2334 AACCTGAAGCGGCTAACACAGACCCACAGCCCCCTACGACACCCCTCTTGGTGTTC 2393
Qy 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
Db 2394 GATATGAGGCGACGGCTCCGACGGCGGTCCCTGAGCTCCCTACCTCCCTCCGCTCC 2453
Qy 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTyrGlySerArgPheLysLysLeuAla 820
Db 2454 GACCAAGACCAAGATTACGATTATCTGAACGAGTGGCGAGCGCTTCAAGAAGCTGGCA 2513
Qy 821 AsnMetTyrGlyGlyGlyGluAspAsp 829
Db 2514 GACATGTACGTGGCGGAGGACGAC 2540
RESULT 7
ABZ24736
ID ABZ24736 standard; cDNA; 3171 BP.
XX
AC ABZ24736;
XX
DT 07-APR-2003 (first entry)
XX
DE Human P-cadherin cDNA.
XX
KW P-cadherin; human; colon cancer; colorectal cancer; cytostatic;
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
CDS 54..2543
FT /*tag= a
FT /product= "Human P-cadherin"
XX
FN WO200297395-A2.
XX
PD 05-DEC-2002.
XX
PF 31-MAY-2002; 2002WO-US017109.
XX
PR 31-MAY-2001; 2001US-0294225P.
XX
PA (CHIR) CHIRON CORP.
XX
PI Reinhard C, Klinger J, Jefferson AB, Escobedo J, Randazo F;
PI Winter J, Goodson R;
XX
DR WPI; 2003-140501/13.
DR P-PSDE; ABP58357.
PT Inhibiting migration and proliferation of P-cadherin expressing cancer
PT for treating cancer, especially digestive cancer, characterized by
PT overexpression of P-cadherin, involves administering a P-cadherin
PT antagonist.
XX
PS Disclosure; Page 11-13; 129pp; English.
XX
CC The present sequence is the nucleic acid sequence for human P-cadherin
CC (placental cadherin), a calcium-dependent cellular adhesion protein. The
CC invention provides methods of treating or diagnosing cancers involving P-
CC cadherin expression using ligands that target P-cadherin, especially
CC human anti-P-cadherin antibodies. A claimed method of treating a cancer
CC characterised by the overexpression and/or upregulation of P-cadherin
CC comprises the administration of a P-cadherin antagonist, optionally
CC conjugated to a therapeutic agent. The migration, adhesion and/or
CC proliferation of the cancer is inhibited, and the method is especially
CC useful for treating or preventing a digestive cancer such as colon or
CC colorectal cancer. The antagonists may be an anti-P-cadherin antibody or
CC its fragment, a ribozyme or antisense oligonucleotide. A transgenic

QY 641 HisGlyHisValGluThrCysProGlyProThrTyrGlyGlyGlyPheIleLeuProValleu 660
 Db 1974 CATGCCCAATGTCGAACCTGCCCTGGACCTGGAAAGGAGGTTTCATCTCCCTGTGCTG 2033
 QY 661 GlyAlaValLeuAlaLeuPheLeuLeuLeuValLeuValLeuValLeuValLeuValLeu 680
 Db 2034 GGGGCTGCTGGCTCTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2093
 QY 681 ArgLysIleLysGluProLeuLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
 Db 2094 CGAAAGATCAAGAGCCCTCTACTCCGAGAAGATGACACCCGTCGACCAAGCTTCTTAC 2153
 QY 701 TyrGlyGluGluGlyGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
 Db 2154 TATGCCGAGAGGGGGGTGGCGAAGAGGACGAGGACTATGACATCACCCAGCTCCACCGA 2213
 QY 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleLeuPro 740
 Db 2214 GGTCTGGAGGCGAGCGGAGGTGTTCTCCGCAATGACGTGGCACCAACCATCATCCCG 2273
 QY 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleLeuGlu 760
 Db 2274 ACACCAATGATCCGTCCTAGGCCACCAACCCAGATGAATCGGCACTTTATATTGAG 2333
 QY 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
 Db 2334 AACCTGAAGCGGTAAACACAGACCCACAGCCCGCTACGACACCCCTCTTGTGTTC 2393
 QY 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
 Db 2394 GACTATGAGGGCAGCGGCTCCGACCGCGCTCCCTGAGCTCCCTCACCTCTCCGCTCC 2453
 QY 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGlySerArgPheLysLeuAla 820
 Db 2454 GACCAAGACCAAGATTACGATTATCTGAACGAGTGGGCGCGCTTCAAGAAGCTGGCA 2513
 QY 821 AspMetTyrGlyGlyGlyGluAspAsp 829
 Db 2514 GACATGTACGTGGCGGGGAGGACGAC 2540
 RESULT 8
 ADD14795
 ID ADD14795 standard; cDNA; 3171 BP.
 AC ADD14795;
 XX AC
 XX AC
 DT 01-JAN-2004 (first entry)
 XX DT
 DE Human src biomarker polynucleotide SEQ ID NO:189.
 XX DE
 KW predictor set; protein tyrosine kinase activity modulator;
 KW protein tyrosine kinase pathway; protein tyrosine kinase; cytostatic;
 KW gene therapy; drug sensitivity; genetic profile; cancer; human; gene; ss.
 XX KW
 OS Homo sapiens.
 XX OS
 PN WO2003062395-A2.
 XX PN
 XX WO2003062395-A2.
 XX WO2003062395-A2.
 PD 31-JUL-2003.
 XX PD
 XX 17-JAN-2003; 2003WO-US001981.
 XX 17-JAN-2003; 2003WO-US001981.
 PR 18-JAN-2002; 2002US-0350061P.
 XX PR
 XX (BRIM) BRISTOL-MYERS SQUIBB CO.
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX PA
 XX Huang F, Fairchild CR, Lee FY, Shaw P;
 XX Huang F, Fairchild CR, Lee FY, Shaw P;
 XX WPI; 2003-636735/60.
 DR P-PSDB; ADD14190.
 XX P-PSDB; ADD14190.
 XX New polynucleotides and polypeptides for predicting the activity of
 PT compounds that interact with protein tyrosine kinases and/or protein

PT tyrosine kinase pathways.
 XX Claim 2; SEQ ID NO 189; 139pp; English.
 XX
 CC The present invention describes a predictor set comprising a plurality of
 CC polynucleotides or polypeptides whose expression pattern is predictive of
 CC the response of cells to treatment with a compound that modulates protein
 CC tyrosine kinase activity or members of the protein tyrosine kinase
 CC pathway. Also described: (1) predicting whether a compound is capable of
 CC modulating the activity of cells, comprising obtaining a sample of cells,
 CC determining whether the cells express a plurality of markers, and
 CC correlating the expression of the markers to the compound's ability to
 CC modulate the activity of the cells; (2) a plurality of cell lines for
 CC identifying polynucleotides and polypeptides whose expression levels
 CC correlate with compound sensitivity or resistance of cells associated
 CC with a disease state; and (3) identifying polynucleotides and
 CC polypeptides that predict compound sensitivity or resistance of cells
 CC associated with a disease state, comprising subjecting the plurality of
 CC cell lines to one or more compounds, analysing the expression pattern of
 CC a microarray of polynucleotides or polypeptides, and selecting
 CC polynucleotides or polypeptides that predict the sensitivity or
 CC resistance of cells associated with a disease state by using the
 CC expression pattern of the microarray. The polynucleotides and
 CC polypeptides have cytostatic activities, and can be used in gene therapy.
 CC The polynucleotides and polypeptides are useful in predicting the
 CC activity of compounds that interact with protein tyrosine kinases and/or
 CC protein tyrosine kinase pathways. These may be used in determining drug
 CC sensitivity in patients to allow the development of individualized
 CC genetic profiles which aid in treating diseases and disorders (e.g.
 CC cancer) based on patient response at a molecular level. The present
 CC sequence is used in the exemplification of the present invention.
 XX
 SQ Sequence 3171 BP; 740 A; 903 C; 864 G; 664 T; 0 U; 0 Other;
 Alignment Scores:
 Pred No.: 0 Length: 3171
 Score: 829.00 Matches: 829
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 9 Gaps: 0
 US-09-916-849A-1 (1-829) x ADD14795 (1-3171)
 QY 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuLeuValCysTrpLeuGln 20
 Db 54 ATGGGGCTCCCTCGTGGACCTCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 113
 QY 21 CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
 Db 114 TGGCGGCTCCGAGCGCTCCGCGCGCTCTTTCAGGGAGGCTGAGTGACCTTGGAGCG 173
 QY 41 GlyGlyAlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
 Db 174 GGAGGCGGAGCAGGAGCGCGGCGCGCTGGGAAAGTATTTCATGGGCTGCGCTGGG 233
 QY 61 GlnGluProAlaLeuPheSerThrAspAsnAspAspPheThrValArgAsnGlyGluThr 80
 Db 234 CAAGAGCAGCTCTGTATTAGCATTGATGATGATGATGATGATGATGATGATGATGATG 293
 QY 81 ValGlnGluArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
 Db 294 GTCCAGGAAAGAGTCACTCACTGAAGAAAGAAATCCATTGAAGATCTTCCATCAACGT 353
 QY 101 IleLeuArgArgHisLysArgAspTrpValValAlaProIleSerValProGluAsnGly 120
 Db 354 ATCTTACGAGACACAAAGAGAGATTGGTGGTGTCTTCAATATCTGTCCCTGAAATGGC 413
 QY 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgThrLys 140
 Db 414 AAGGTCCTCTCCCGCAGAGACTGATCAGCTCAAGTCTAATTAAGATAGACACCAAG 473
 QY 141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProGluGlyValPheAlaVal 160

Db 131 TCGCGGGCTCCGAGCCGTGCGGGGTCTTCAAGGAGGCTGAAGTGAACCTTGAGGCG 190
 Qy 41 GlyGlyAlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
 Db 191 GGAGCGCGGACGAGAGCCCGCCAGCGCTGGGGAAGATTTCATGGCTGCCCTGGG 250
 Qy 61 GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr 80
 Db 251 CAAGAGCCAGCTCTGTTAGCACTGATAATGATGACTTCACCTGTCGGAATGGCGAGACA 310
 Qy 81 ValGlnGluArgArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
 Db 311 GTCCAGGAAAGAGTCACTGAAGAAAGGAATCCATTGAAGATCTTCCCATCCAAACGT 370
 Qy 101 IleLeuArgArgHisLysArgAspTyrValAlaLapIleSerValProGluAsnGly 120
 Db 371 ATCTTACGAGACACAAGAGATTGGGTGGTGTCTCCAAATATCTGTCCCTGAAAATGGC 430
 Qy 121 LysGlyProPheProGluArgLeuAsnGlnLysSerAsnLysAspArgAspThrLys 140
 Db 431 AAGGTCCTCTTCCCCCAGAGACTGAATCAGCTCAAGTCTAATGAAGATAGACACCAAG 490
 Qy 141 IlePheTyrSerIleThrGlyProGlyValAlaAspSerProGluGlyValPheAlaVal 160
 Db 491 ATTTTCTACAGCATCACGGGCGCGGGGCGAGACAGCCCCCTGAGGGTGTTCTTCGTGTA 550
 Qy 161 GluLysGluThrGlyTyrIleLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys 180
 Db 551 GAGAAGGAGACAGGCTGCTGTTGTTGAATAAGCCACTGGACCCGGAGGAGATTGCCAAG 610
 Qy 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
 Db 611 TATGAGCTCTTTGGCCACGCTGTGTACAGAGATGGTCTCAGTGGAGGACCCCATGAAC 670
 Qy 201 IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
 Db 671 ATCTCCATCATCGTACCGACAGCAAGTACCACCAAGCCCAAGTTTACCCAGACACCTTC 730
 Qy 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
 Db 731 CGAGGAGTGTCTTAGGGGAGTCTACACAGTACTTCTGTGATGCAGGTGACAGCCACG 790
 Qy 241 AspGluAspAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln 260
 Db 791 GATGAGGATGATGCATCTACCTACAAATGGGTGGTGTGTTACTCCATCCATAGCCAA 850
 Qy 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
 Db 851 GAACCAAGGACCCACACGACCTCATGTTCCACATTCACCGAGACAGGACCATCAGC 910
 Qy 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
 Db 911 GTCATCTCCAGTGGCTGGACCGGAAAGTCCCTGAGTACACACTGACCATCCAGGCC 970
 Qy 301 ThrAspMetAspGlyAspGlySerThrThrThrAlaValAlaValAlaValGluIleLeuAsp 320
 Db 971 ACAGACATGGATGGGACGGCTCCACCAACGCGCAGTGGCAGTAGTGGAGATCCTTGAT 1030
 Qy 321 AlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
 Db 1031 GCCAATGACATGCTCCCATGTTTGAOCCCGCAGAAAGTACGAGGCCCATGTGCTCGAGAAT 1090
 Qy 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
 Db 1091 GCAGTGGCCCATGAGTGCAGAGGCTGACGCTCACTGATCTGGACGCCCCCACTCACCA 1150
 Qy 361 AlaTyrArgAlaThrTyrLeuIleMetGlyLysAspAspGlyAspHisPheThrIleThr 380
 Db 1151 GCGTGGCTGTCACCTACTATCATGGCGGTGACGACGGGGACCAATTTTACCATCACC 1210
 Qy 381 ThrHisProGluSerAsnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
 Db 1211 ACCACCTTGAGGACACAGGCAATCTTGACAAACAGGAGGGTTTGGATTTTGGAGGCC 1270

Qy 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
 Db 1271 AAAAACAGACACACCTGTACGTTGAAGTGACCAACAGAGCCCTTTTGTGCTGAAGCTC 1330
 Qy 421 ProThrSerThrAlaThrIleValHisValGluAspValAsnGluAlaProValPhe 440
 Db 1331 CCAACCTCCACAGCCACCATAGTGGTCCACGTGGAGGATGTGAATGAGGCACCTGTGTTT 1390
 Qy 441 ValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
 Db 1391 GTCCCACTCCCAAAGTCGTGAGGTCACAGAGGACATCCCCACTGGGAGCCCTGTGTGT 1450
 Qy 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
 Db 1451 GTCTACACTGCAGAGACCCCTGACAAAGGAGAAATCAAAGATCAGCTACCGACTCCTGAGA 1510
 Qy 481 AspProAlaGlyTyrIleAlaMetAspProAspSerGlyClnValThrAlaValGlyThr 500
 Db 1511 GACCCACGAGGGTGGGTAGCCATGSCACCCAGACAGTGGGCGAGGTACAGCTGTGGGCACC 1570
 Qy 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
 Db 1571 CTGACCGTGAGGATGAGCAGTTTGTGAGGAACACATCTATGAAGTCACTGTCTTGGCC 1630
 Qy 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAsp 540
 Db 1631 ATGACAAATGGAAGCCCTCCACCACTGGCACCGGAACCCCTCTGCTAAACACTGATTGAT 1690
 Qy 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
 Db 1691 GTCAATGACCATGGCCAGTCCCTGAGCCCGCTCAGATCACCATCTGCAACCAAGCCCT 1750
 Qy 561 ValArgHis-ValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
 Db 1751 GTGCGCCA-GGTGCTGAACATCACGGACAAAGGACCTGTCTCCCAACACCTCCCTTTCCA 1809
 Qy 580 nAlaGlnLeuThrAspAspSerAspIleTyrThrAlaGluValAsnGluGluLys 600
 Db 1810 GGCCAGCTCACAGATGACTCAGACATCTACTGCGCGCAGAGGTCAACGAGGAAGGTGA 1869
 Qy 600 pThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSe 620
 Db 1870 CACAGTGGTCTTGCTCCCTGAAGAAGTTCCTGAAGCAGGATACATATGACGTGCACCTTC 1929
 Qy 620 rLeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
 Db 1930 TCTGTCTGACCATGGCAACAAAGAGAGAGCTGACGGTGATCAGGGCCACTGTGTGCGACTG 1989
 Qy 640 sHisGlyHisValGluThrCysProGlyProTyrLysGlyGlyPheIleLeuProValle 660
 Db 1990 CCATGGCCATGTCGAAACCTGCCCTGGAAAGGAGGTTTCATCTCTCCCTGTGCT 2049
 Qy 660 uGlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysLys 680
 Db 2050 GGGGGCTGCTCGCTCTGCTGTCTCTCTGCTGGTGTCTGCTTTCTGGTGAGAAAGAA 2109
 Qy 680 sArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPhe 700
 Db 2110 GCGGAAGATCAGAGGCCCTCTCTACTCCAGAAAGATGACACCCGCGTGAACGCTTCTTA 2169
 Qy 700 rTyrGlyGluGluGlyGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisAr 720
 Db 2170 CTATGCGGAAGAGGGGGTGGCGAAGAGGACCCAGGACTATGACATCACCCAGCTCCACCG 2229
 Qy 720 gGlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIleLe 740
 Db 2230 AGGTCTGGAGGCCAGGCCGAGGTGTTCTCCGCAATGACGTGGCAACCAACCATCATCC 2289
 Qy 740 oThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGln 760
 Db 2290 GACACCCATGTACCGTCTCGGCCAGCCAAACCCAGATGAAATCGGCAACTTTATATTTGA 2349

QY 760 uAsnLeuLysAlaAAsnThrAspProThrAlaProProTrpAspThrLeuLeuValPh 780
 DB 2350 GAACCTGAAGCGCGTAAACACAGACCCACAGCCCGCCCTACGACACCCCTCTTGGTGT 2409
 QY 780 eAspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerAlaSe 800
 DB 2410 CGACTATGAGGCGAGCGGCTCGACGCGCGTCCCTGAGCTCCCTCACCTCTCGGCTC 2459
 QY 800 rAspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGlySerArgPheLysLysLeuAl 820
 DB 2470 CGACCAAGACCAAGATTACGATTATCTGAACGAGTGGGCGAGCGCTTCAAGAAAGCTGGC 2529
 QY 820 aAspMetTyrGlyGlyGlyGluAspAsp 829
 DB 2530 AGACATGTACGTTGCGGGGAGGACGAC 2557

RESULT 11
 ACC72821
 ID ACC72821 standard; cDNA; 3205 BP.
 XX
 AC ACC72821;
 XX
 DT 09-JUL-2003 (first entry)
 XX
 DE Human cancer related protein encoding cDNA SEQ ID NO:159.
 XX
 KW Human; cancer; diagnosis; screening; modulator; leukaemia; ischaemia;
 KW heart disease; atherosclerosis; endometriosis; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO2003025138-A2.
 XX
 PD 27-MAR-2003.
 XX
 PF 17-SEP-2002; 2002WO-US029560.
 XX
 PR 17-SEP-2001; 2001US-0323469P.
 PR 20-SEP-2001; 2001US-0323887P.
 PR 13-NOV-2001; 2001US-0350666P.
 PR 08-FEB-2002; 2002US-0355145P.
 PR 08-FEB-2002; 2002US-0355257P.
 PR 12-APR-2002; 2002US-0372246P.
 XX
 PA (EOSB-) EOS BIOTECHNOLOGY INC.
 XX
 PI Afar D, Aziz N, Gish KC, Hevezi PA, Mack DH, Wilson KE;
 PI Zlotnik A;
 XX
 DR WPI; 2003-354600/33.
 DR P-PSDB; ABR58670.
 XX
 PT New genes that are up-regulated or down-regulated in cancers, useful as
 PT markers for diagnosing e.g. cancer, ischemia or heart diseases, or as
 PT therapeutic targets for screening drugs for treating these diseases.
 XX
 PS Claim 8; Page 715; 767pp; English.

The present invention describes an isolated nucleic acid molecule, which comprises the sequence of any of the genes that are up-regulated or down-regulated in specific cancers (e.g. about 1031 genes up-regulated in acute lymphocytic leukemia). ACC72641 to ACC72860 represent cancer related gene nucleotide sequences which encode the proteins given in ABR58521 to ABR58709. Also described: (1) determining the presence or absence of a pathological cell in a patient; (2) an expression vector comprising a nucleic acid molecule described above; (3) a host cell comprising the vector; (4) an isolated polypeptide, which is encoded by the nucleic acid; (5) an antibody that specifically binds the polypeptide of (4); (6) specifically targeting a compound to a pathological cell in a patient by administering to the patient the antibody above; and (7) a drug screening assay. The nucleic acid is useful as diagnostic markers or therapeutic targets. In particular, the nucleic acid is useful for diagnosing a pathology, e.g. cancer (e.g. cancer of the bone marrow,

CC bladder, brain, breast, cervix, colon/rectum, kidney, lung, ovary,
 CC pancreas, prostate, skin and uterus), wounds, ischaemia, heart diseases,
 CC atherosclerosis and endometriosis. The nucleic acid is also useful in
 CC drug screening, particularly for identifying agents for treating these
 CC pathologies
 XX
 SQ Sequence 3205 BP; 751 A; 909 C; 875 G; 670 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 0 Length: 3205
 Score: 728.00 Matches: 828
 Percent Similarity: 99.76% Conservative: 0
 Best Local Similarity: 99.76% Mismatches: 1
 Query Match: 87.82% Indels: 2
 DB: 7 Gaps: 0

US-09-916-849A-1 (1-829) x ACC72821 (1-3205)

QY 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuLeuValCysTrpLeuGln 20
 DB 71 ATGGGGCTCCCTCGTGGACCTCTCGCGTCTCTCCCTCTCCAGGTTTGGCTGGCAG 130
 QY 21 CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
 DB 131 TGGCGGGCTCCGAGCGGTCGCGGGGCTTCGAGGAGGCTGAAGTGACCTTGGAGGCG 190
 QY 41 GlyGlyValaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
 DB 191 GGAGGCGCGAGCAGGAGCCCGCCAGCGCTGGGGAAAGTATTTCATGGCTGCCCTGGG 250
 QY 61 GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr 80
 DB 251 CAAGAGCCAGCTCTGTTTAGCCTGATAATGATGACTTCACCTGCGGAATGGCAGACA 310
 QY 81 ValGlnGluArgArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
 DB 311 GTCCAGGAAGAAGGTCACTGAAGGAAGAAGATCCATTCAAGATCTTCCATCCAAAGCT 370
 QY 101 IleLeuArgArgHisLysArgAspTrpValValAlaProIleSerValProGluAsnGly 120
 DB 371 ATCTTACGAAGACACAGAAGAGATTGGTGGTGTCTCAATATCTGCTCCCTGAATGGC 430
 QY 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
 DB 431 AAGGTTCCTTCCCGCAGAGACTGAATCAGCTCAAGTCTAATAAGATAGACACCAAG 490
 QY 141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProGluGlyValPheAlaVal 160
 DB 491 ATTTTCTACAGCATCACGGGGCGGGGCGAGACAGCCCGCCCTGAGGGGTCTTCTGCTGTA 550
 QY 161 GluLysGluThrGlyTrpLeuLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys 180
 DB 551 GAGAAGGAGACAGCGCTGGTGTGTTGAATAGCCACTGACCGGAGGAGAGATTGCCAAG 610
 QY 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
 DB 611 TATGAGCTCTTTGGCCACGCTGTGTGAGAGAATGGTCCCTCAGTGAGGAGACCCCATGAAC 670
 QY 201 IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
 DB 671 ATCTCCATCATCTGTAGCCAGCAGACCAAGATACCAAGCCCAAGTTTACCAGGACACCTTC 730
 QY 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
 DB 731 CGAGGGAGTGTCTTAGGGAGTCTTACAGGTACTTCTGTGATGTCAGGTGACGACCAAG 790
 QY 241 AspGluAspAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln 260
 DB 791 GATGAGATGATCCCATCATACCTCAATGGGGTGGTGTCTTACTTCCATCCATGAGCCAA 850
 QY 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
 DB 851 GATCCAAAGAGACCCACACGACCTCATGTTCCATTCACCGGAGCAGACGACCATCAGC 910

QY 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
DB 911 GTCACTCTCCAGTGGCTGGACCGGAAAAAGTCCCTGAGTACACACTACCATCCAGGCC 970
QY 301 ThrAspMetAspGlyAspGlySerThrThrThrAlaValAlaValGluIleLeuAsp 320
DB 971 ACAGACATGGATGGGAGCGCTCCACCACCGCAGTGGCAGTAGTGAGATCCTTGAT 1030
QY 321 AlaAspAspAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
DB 1031 GCCAATGACAATGCTCCCATGTTTGAACCCCGAGAAAGTACGAGGCCCATGTGCTGAGAA 1090
QY 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
DB 1091 GCAGTGGGCGCATGAGGTGAGAGCTGACGGTCACTGATCGGACGCCCCCAACTCACCA 1150
QY 361 AlaTyrArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
DB 1151 GGTGCGGTGGCCACTCTATCATCTATCATGGCGGTGACGAGCGGGACCATTTTACCATCAC 1210
QY 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
DB 1211 ACCCACTCGAGCAACAGCGGCATCTGCAACACGAGAGGGTGTGATTTTGGAGGCC 1270
QY 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
DB 1271 AAAACCAAGCACACCTGTACGTTGAAGTGAACACGAGGCCCTTTGTGCTGAGGCTC 1330
QY 421 ProThrSerThrAlaThrIleValHisValGluAspValAsnGluAlaProValPhe 440
DB 1331 CCAACTCCACAGCCACCATAGTGTGTCACGTGGAGGATGTGAATGAGGCACCTGTGTGT 1390
QY 441 ValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
DB 1391 GTCCCACTCCCTCAAAAGTCTGTAGTCCAGAGGGCATCCCACTGGGGAGCCCTGTGTGT 1450
QY 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
DB 1451 GTCTACACTGCAGAGAGCCCTGACAGGAGGAATCAAAAGATCAGCTACCGCATCTCGAGA 1510
QY 481 AspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
DB 1511 GACCCAGCAGGGTGGCTAGCATGTGACCCAGACAGATGGGCGAGGTACACAGCTGTGGCACC 1570
QY 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
DB 1571 CTCGACCGTGGAGTGCAGCATTTGTGAGAGAACACATCTATGAGTCACTGGTCTTGCC 1630
QY 521 MetAspAsnGlySerProProThrThrThrGlyThrLeuLeuLeuThrLeuLeuAsp 540
DB 1631 ATGGACAATGGAAGCCCTCCACCACTGGCACGGGAACCTTCTGCTAAACACTGATTGAT 1690
QY 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
DB 1691 GTCAATGACCATGGCCCACTCCCTGAGCCCGTCAGATCACCATCTGCAACCAAGGCCCT 1750
QY 561 ValArgHis-ValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
DB 1751 GTGGCCCA-GGTGCTGAACATACCGACAGGACCTGTCTCCACACCTCCCTCTTCCA 1809
QY 580 nAlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGluLys 600
DB 1810 GGCCCGAGCTCAGATGACTCAGACATCTACTTGACCGCAGAGTCAACGAGGAAGTGA 1869
QY 600 pThrValValLeuSerLeuLysPheLeuLysGlnAspThrTyrAspValHisLeuSe 620
DB 1870 CACAGTGTCTGTCTTCCCTGAGAGAGTCTCTGACGAGGATACATGACGTGCACCTTTC 1929
QY 620 rLeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCy 640
DB 1930 TCTGTCTGACCATGGCAAAAGAGCAGCTGACGGTGTATCAGGGCCCATGTGTGCGCACTG 1989

QY 640 sHisGlyHisValGluThrCysProGlyProTyrLysGlyGlyPheIleLeuProValle 660
DB 1990 CCATGGCCATGTGAAAACCTGCGCTGGACCTGGAAGGGAGGTTTCATCTCCCTGTGCT 2049
QY 660 uGlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuValLeuValArgLysLy 680
DB 2050 GGGGCTGTCTCTGGCTCTGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2109
QY 680 sArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTy 700
DB 2110 GCGGAAGATCAAGGAGGCCCTCTACTCCAGAAAGATGACACCCGTGACACACTCTCTA 2169
QY 700 rTyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisAr 720
DB 2170 CTATGGCGAAGAGGGGGGTGGCGAAGGAGGACCAAGGACTATGACATCACCAGCTCCACCG 2229
QY 720 gGlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePr 740
DB 2230 AGTCTGGAGGCCAGCGCGGAGGTGTTCTCCGCANTGACGTGGCACCACCATCATCCC 2289
QY 740 oThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGl 760
DB 2290 GACACCATGTATCCCTCTCGCCAGCAACCCAGATGAATCGGAACCTTTATAATTGA 2349
QY 760 uAsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPh 780
DB 2350 GAACCTGAAGCGCGCTTAACACAGACCCCGCCAGCCCGCTTACGACACCTCTCTGTGT 2409
QY 780 eAspTyrGluGlySerGlySerAspAlaLaserLeuSerSerLeuThrSerSerAlase 800
DB 2410 CGACTATAGGGCGAGCGGTCCGACCGCGTCCCTGAGTCCCTCACCTCTCCGCTC 2469
QY 800 rAspGlnAspGlnAspTyrAspTyrLeuAsnGluTyrGlySerArgPheLysLeuAl 820
DB 2470 CGACCAAGACCAAGATTACGATTATCTGAACGAGTGGGCGAGCCGCTTCAAGAAGCTGCC 2529
QY 820 aAspMetTyrGlyGlyGlyGluAspAsp 829
DB 2530 AGACATGTACGTGGCGGGAGGAGCGAC 2557
RESULT 12
ABX76155
ID ABX76155 standard; DNA; 3205 BP.
XX AC ABX76155;
XX DT 02-APR-2003 (first entry)
XX DE Lung cancer-associated polynucleotide #27.
XX KW Lung cancer-associated polynucleotide; gene; ds; cytostatic; emphysema;
XX KW antiinflammatory; antiaschmatic; non-small cell lung cancer; atelectasis;
XX KW small cell lung cancer; benign lesion; precancerous lesion; bronchitis;
XX KW chronic obstructive pulmonary disease; hypersensitivity pneumonitis;
XX KW interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.
OS Unidentified.
XX WO200286443-A2.
XX PN 31-OCT-2002.
XX PD 18-APR-2002; 2002WO-US012476.
XX PF 18-APR-2001; 2001US-0284770P.
XX PR 10-MAY-2001; 2001US-0290492P.
XX PR 09-NOV-2001; 2001US-0339245P.
XX PR 13-NOV-2001; 2001US-0350666P.
XX PR 29-NOV-2001; 2001US-0334370P.
XX PR 12-APR-2002; 2002US-0372246P.
XX PA (BOSB-) BOS BIOTECHNOLOGY INC.

Qy 21 CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
Db 131 TCGGCGGCTCCGAGCGGTGCGGGGGTCTTCAGGGAGGTGAAGTGAACCTTGGAGGG 190
Qy 41 GlyIyAlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
Db 191 GGAGGCGCGAGCAGAGCGCGCGAGCGCTGGGGAAGTATTATCGGGCTGCCCTGGG 250
Qy 61 GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr 80
Db 251 CAAGAGCCAGCTCTGTTAGCACTGATATGATGACTTCACTGTCGGAAATGGCGAGACA 310
Qy 81 ValGlnGluArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
Db 311 GTCCAGGAAGAAGGTCACTGAAGGAAGGAATCCATTGAAGATCTTCCCATCCAAACGT 370
Qy 101 IleLeuArgArgHisLysArgAspThrValAlaIleProIleSerValProGluAsnGly 120
Db 371 ATCTTACGAAGACACAAGAGAGATTGGGTGGTGTCTCCAAATATCTGTCCTGAAAAATGGC 430
Qy 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgThrLys 140
Db 431 AAGGGTCCCTTCCCCAGAGACTGAATCAGCTCAAGTCTAATAAGATAGAGACACCAAG 490
Qy 141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProGluGlyValPheAlaVal 160
Db 491 ATTTTCTACAGCATCACGGGGCGGGGCGAGACAGCCCCCTGAGGGTGTCTTCGCTGTA 550
Qy 161 GluLysGluThrGlyTyrLeuLeuLeuLysProLeuAspArgGluGluIleAlaLys 180
Db 551 GAGAGGAGACAGGCTGGTGTGTTGTAATAGCCACTGGACCGGAGAGATTGCCAAG 610
Qy 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
Db 611 TATGAGCTCTTTGGCCACGCTGTGTGAGAGATGGTGTCTCAGTGGAGGACCCCATGAAC 670
Qy 201 IleSerIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
Db 671 ATCTCCATCATCGTACCGACCAAGATGACCAAGCCCAAGTTTACCAGGACACCTTC 730
Qy 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
Db 731 CGAGGGAGTGTCTAGAGGAGTCTTACCGAGTACTTCTGTGATCGAGGTGACAGCCACG 790
Qy 241 AspGluAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln 260
Db 791 GATGAGGATGATGCCATCTACACCTTACAAATGGGGTGGTGTCTTACTCCATCCATAGCCAA 850
Qy 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
Db 851 GAACCAAGGACCCACAGCCTCATGTTCACCATTCACCGGAGCACAGGCACCATCAGC 910
Qy 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
Db 911 GTCATCTCCAGTGGCTGTGACCGGGAAGTCCCTGAGTACACACTGACCATCCAGGCC 970
Qy 301 ThrAspMetAspGlyAspGlySerThrThrAlaValAlaValGluIleLeuAsp 320
Db 971 ACAGACATGGATGGGACCGGCTCCACCACCGAGGAGTGGCAGTGTAGTGGAGATCCTTGAT 1030
Qy 321 AlaAsnAspAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
Db 1031 GCCAATGCAATGCTCCCATGTTTACCACCAAGAGTACGAGGCCCATGTGCTGAGAAAT 1090
Qy 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
Db 1091 CGAGTGGGCGCATGAGGTGAGAGGTGACGGTCACTGTATCTGGACGCCCCCACTCACA 1150
Qy 361 AlaTyrArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
Db 1151 GCGTGGCGTGCACCTTACCTTATCATGGCGGTGACGCGGGACCAATTTACCATCACC 1210

Qy 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgGlyGlyLeuAspPheGluAla 400
Db 1211 ACCCACTTGGAGAGCAACAGGGCATCTTCACAACCCAGGAAGGTTTGGATTTTGGAGGCC 1270
Qy 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db 1271 AAAAACCCAGCACACCTGTACGTTGAAGTGAACACAGAGGCCCTTTGTGCTGAAGCTC 1330
Qy 421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
Db 1331 CCAACCTCCACAGCCACCATAGTGTGTCCACGTGGAGATGTGAATGAGGCACCTGTGTTT 1390
Qy 441 ValProProSerLysValValGluValGlnGlyIleProThrGlyGluProValCys 460
Db 1391 GTCCCCACCTTCCAAAGTGTGTGAGTCCAGGAGGGCATCCCCACTGGGAGGCTGTGTGT 1450
Qy 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db 1451 GTCTACACTGCAGAGACCCCTGACAGGAGATCAAAAGATCAGCTACCGCATCCTGAGA 1510
Qy 481 AspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db 1511 GACCCAGCAGGTGGTGTAGCCATGGACCCAGACAGTGGGCGAGTCCACAGCTGTGGCACC 1570
Qy 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db 1571 CTGACCGTGGAGTGAAGCAGTTTGTGAGAACACACATCTATGAAGTCACTGTCTGGCC 1630
Qy 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAsp 540
Db 1631 ATGGACAATGGAAGCCCTCCACCACTGGCAGCGGAAACCTTCTGTAAACACTGATTGAT 1690
Qy 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 1691 GTCAATGACCATGGCCCGAGTCCCTGAGCCCCGTGAGATCACCATCTGCAACCAAGCCCT 1750
Qy 561 ValArgHis-ValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
Db 1751 GTGGGCCA-GGTGCTGAACATCACGACGAAGGACCTGTCTCCCAACACCTCCCTTCCA 1809
Qy 580 nAlaGlnLeuThrAspAspSerAspIleTyrThrAlaGluValAsnGluGluGlyAs 600
Db 1810 GGCCAGCTCACAGATGACTCAGACATCTACTGAGCGGAGAGGTCAACGAGGAAGGTGA 1869
Qy 600 pThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSe 620
Db 1870 CACAGTGGTCTTGTCCCTGAGAGTTCCTGAAGCAGGATACATATGACGTGCACCTTTC 1929
Qy 620 rLeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCy 640
Db 1930 TCTGTCTGACCATGGCAACAAGAGCAGCTGACGTGATCAGGGCCACTGTGTGGAGTGG 1989
Qy 640 sHisGlyHisValGluThrCysProGlyProTyrLysGlyGlyPheIleLeuProValle 660
Db 1990 CCATGGCCATGTCCAAACCTGCTCCCTGAGACCTTGGAGAGGAGGTTTCATCTCTCCCTGTGCT 2049
Qy 660 uGlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValLeuValArgLysLy 680
Db 2050 GGGGGCTGTCTGGCTGTCT 2109
Qy 680 sArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTy 700
Db 2110 GCGGAAGATCAAGAGGCCCTCTCTACTCCCAAGATGACACCCGCTGACACAGTCTTCTA 2169
Qy 700 rTyrGlyGluGluGlyGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisAr 720
Db 2170 CTATGGCAGAGAGGGGGTGGCGAAGAGGACAGGACTATGACATCACCCAGCTCCACCG 2229
Qy 720 sGlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePr 740
Db 2230 AGGTCTGGAGGCCAGCGGAGGTGTTCTTCGCAATGACGTGGGCAACCAACCATATCCC 2289
Qy 740 oThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGl 760

QY 301 ThrAspMetAspGlyAspGlySerThrThrAlaValAlaValValGluLeuLeuAsp 320
 Db 990 ACAGACATGGATGGGACGGCTCCACCACCGAGTGGCAGTAGTGGAGATCCTTGAT 1049
 QY 321 AlaAsnAspAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
 Db 1050 GCCAATGCAATGCTCCCATGTTTACCCCCAGAAAGTACGAGGCCCATGTGCTGAGAA 1109
 QY 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
 Db 1110 GCAGTGGGCCATGAGGTGCAGAGGTGACGGTCACTGATCTGGACGCCCACTCACCA 1169
 QY 361 AlaTyrArgAlaThrTyrLeuLeuMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
 Db 1170 GCGTGGCGTGCACCTTACCTTATCATGGCGGTGACGCGGAGCCATTTTACCATCAC 1229
 QY 381 ThrHisProGluSerAsnGlnGlyLeuThrThrArgLysGlyLeuAspPheGluAla 400
 Db 1230 ACCACCTTGAGAGCAACCCAGGGCATCTGACACACAGGAGGTTCGATTTTGGAGCC 1289
 QY 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
 Db 1290 AAAAACCCAGCACACCTGTACCTTGAAGTGACCAACGAGGCCCTTTTGTGCTGAAGCTC 1349
 QY 421 ProThrSerThrAlaThrIleValHisValGluAspValAsnGluAlaProValPhe 440
 Db 1350 CCAACCTCCACAGCCACCATAGTGTCCACGTGGAGGATGTGAATGAGGCACTGTGT 1409
 QY 441 ValProProSerLysValValGluValGlnGlyIleProThrGlyGluProValCys 460
 Db 1410 GTCCCACTCCAAAGTCTGTGAGGTCCAGAGGGCATCCCCACTGGGAGCGCTGTGT 1469
 QY 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleuArg 480
 Db 1470 GTCTACACTGCAGAGACCTCAGCAGGAGATCAAAAGATCAGCTACCGATCTCTGAGA 1529
 QY 481 AspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
 Db 1530 GACCCAGCGGGTGGTACCCATGACCCAGACAGTGGGAGTCCACAGCTGTGGGCACC 1589
 QY 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
 Db 1590 CTCGACCGTGAGATGAGCAGTTTGTGAGGAAACAATCTATGAAGTCACTGCTGTGCCC 1649
 QY 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuLeuAsp 540
 Db 1650 ATGGACAAATGGAAAGCCCTCCACCACTGGCAGCGGAAACCTTCTCTAACACTGATGAT 1709
 QY 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
 Db 1710 GTCAACGACCATGGCCCATCTCTCCAGCCCGCTCAGATCACCATCTGCAACCAAGCCCT 1769
 QY 561 ValArgHis - Val - IeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheG 580
 Db 1770 GTGCGCCA - GGTGGCTGGAATCAGGCAAGACCTGTCTCCACACCTCCCTTTCC 1828
 QY 580 IuAlaGlnLeuThrAspAspSerAspIleTyrTriThrAlaGluValAsnGluGlyAla 600
 Db 1829 AGGCCAGCTCAGATGACTCAGACATCTACTGGACGCGCAGAGTCAACGAGGAAGGTG 1888
 QY 600 spThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuS 620
 Db 1889 ACACAGTGTCTTGTCTCCCTGAAGAAGTTCTCTGAACGAGGATACATATGACGTGCACCTT 1948
 QY 620 ertLeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspC 640
 Db 1949 CTCTGTCTGACCATGTCACAAAGAGCAGCTACGGTGTATCAGGCCACTGTGTGGACT 2008
 QY 640 yHisGlyHisValGluThrCysProGlyProTyrLysGlyGlyPheIleLeuProValL 660
 Db 2009 GCCATGCCATGTCGAAACCTGCGCTGGACCCCTGGAAGGAGGTTCATCTCCCTCTGTC 2068
 QY 660 euGlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysL 680

Db 2069 TGGGGGCTGCTCCGGCTCTGCTGTTCTCTCTGCTGCTGCTGTTTGTGGTGAAGA 2128
 QY 680 ySArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheT 700
 Db 2129 AGCGGAAGATCAGAGAGCCCTCTACTCCAGAGATGACACCCGTCGACACGCTCTTCT 2188
 QY 700 yTyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisA 720
 Db 2189 ACTATGCGGAAGAGGGGGTGGCGAAGAGCAGAGCATGACATCACCCAGCTCCACC 2248
 QY 720 rgGlyLeuGluAlaAspProGluValValLeuArgAsnAspValAlaProThrIleIleP 740
 Db 2249 GAGGTGTGGGGCCAGCCGCGAGGTGTTCTCCGCAATGACGTGGCACCACCATCATCC 2308
 QY 740 roThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleG 760
 Db 2309 CGACACCATGTACCTGCTAGGCCAGCCACCCAGATGAAATCGGCACTTTATAATTG 2368
 QY 760 IuAsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValP 780
 Db 2369 AGACCTTGAGGCGGTTCACACAGACCCACAGCCCGCCCTACGACACCTCTTGGTGT 2428
 QY 780 heAspTyrGluGlySerGlySerAspAlaAlaSerLeuSerLeuThrSerSerAlaS 800
 Db 2429 TCGACTATGAGGCGCAGCGCTCCGACCGCGCTCCCTGAGTCCCTCACCTCTCCGCT 2488
 QY 800 ertAspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGlySerArgPheLysLeuA 820
 Db 2489 CCACCAAGACCAAGATTACGATTATCTGAACGAGTGGGCGCCCTTCAGAAAGCTGG 2548
 QY 820 IuAspMetTyrGlyGlyGlyGluAspAsp 829
 Db 2549 CAGACATGTACGTGGCGGGAGGACGAC 2577

RESULT 15

AAH81822

ID AAH81822 standard; DNA; 306 BP.

XX AAH81822;

XX 21-SEP-2001 (first entry)

DE Human differential transcription-associated cDNA SEQ ID 331.

XX Differential transcription; human; rat; tumour cell; cytostatic;

KW Ras modulator; Class II tumour suppressor gene; gene therapy; ss.

OS Homo sapiens.

XX WC200157058-A2.

XX 09-AUG-2001.

XX 31-JAN-2001; 2001WO-EP001003.

XX 31-JAN-2000; 2000DE-01004102.

XX (META-) METAGEN GES GENOMFORSCHUNG MBH.

XX Rosenthal A, Hinzmann B, Schaefer R, Zuber J, Tchernitsa O;

XX Grips M, Hellriegel M, Schmitz A, Sers C;

XX WPI; 2001-483415/52.

XX Nucleic acids differentially expressed between tumor and normal cells,
 useful for diagnosis or therapy of tumors and for screening active
 agents.

XX Disclosure; Page 475; 579pp; German.

XX This invention describes a nucleic acid (I) with differential expression
 between tumour and normal cells and which has cytostatic activity. (I)

CC work as modulators of Ras activity by inducing expression of tumour
CC suppressor genes. (1), and polypeptides encoded by them, are useful as
CC targets for diagnosis or therapy and in screening to determine the
CC effects of an active compound (potential pharmaceutical) on a cell line,
CC particularly for diagnosis and treatment of tumors, especially by
CC modulating expression of (1) (by gene therapy, antisense RNA or ribozyme
CC methods) or by modulating the amount and/or location of (1)-encoded
CC polypeptides (by administration of the polypeptide or its activator,
CC antibody (optionally as a conjugate) or inhibitor). The method allows
CC identification of many class II tumour suppressor genes (i.e. genes that
CC are not primary targets for tumour-initiating mutations). AAH81492-
CC AAH82376 represent the human and rat derived nucleic acid fragments
CC described in the method of the invention
XX

SQ Sequence 306 BP; 73 A; 110 C; 73 G; 50 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.55e-86	Length:	306
Score:	102.00	Matches:	102
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	12.30%	Indels:	0
DB:	5	Gaps:	0

US-09-916-849A-1 (1-829) x AAH81822 (1-306)

QY	707	GlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisArgGlyLeuGluAlaArgPro	726
DB	1	GGCGAAGAGGACGAGACTATGACATCACCCAGCTCCACCGAGGTCTGGAGGCCAGGCCG	60
QY	727	GluValValLeuArgAsnAspValAlaProThrIleIleProThrProMetTyrArgPro	746
DB	61	GAGGTGGTTCCTCCGCAATGACGTGGCACCACCATCTCCCGACACCCATGTACCGTCT	120
QY	747	ArgProAlaAsnProAspGluIleGlyAsnPheIleIleGluAsnLeuIleAlaAlaAsn	766
DB	121	CGGCCAGGCCAACCCAGATGAATCGCACTTTAATAGAACCTGAGGCGGCTAAC	180
QY	767	ThrAspProThrAlaProProTyrAspThrLeuValPheAspTyrGluGlySerGly	786
DB	181	ACAGACCCACAGCCCGCCCTACGACACCCCTCTTGGTGTTCGACTATGAGGCGACGGCC	240
QY	787	SerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSerAspGlnAspTyr	806
DB	241	TCCGACGCGCGCTCCCTGAGCTCCCTCAGCTCTCTCGGCTCCGACCAAGACCAAGATTAC	300
QY	807	AspTyr	808
DB	301	GATTAT	306

Search completed: September 23, 2004, 03:59:03
Job time : 855 secs

This Page Blank (uspto)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 23, 2004, 03:29:35 ; Search time 144 Seconds
(without alignments)
3194.821 Million cell updates/sec

Title: US-09-916-849A-1

Perfect score: 829

Sequence: 1 MGLPRGLASLLQLVCWLQ.....NEWGSRFKKLMYGGGDD 829

Scoring table: OLIGO

Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1360453

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters: -DEV=xlh

-O=/cgn2_1/USPTO.spool.h/US0916849/runat.21092004.145836.7341/app.query.fasta.1.967
-DB=Issued_Patents_NA -QFWT=fastap -SUFFIX=oligo.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFWT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US0916849 @CNC 1.1.69 @runat.21092004.145836.7341 -NCPU=6 -ICPU=3
-NO MNAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA.*

1: /cgn2_6/ptodata/2/ina/5A-COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B-COMB.seq.*
3: /cgn2_6/ptodata/2/ina/5A-COMB.seq.*
4: /cgn2_6/ptodata/2/ina/5B-COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS-COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	829	100.0	3171	4	US-09-016-434-1348
2	14	1.7	2808	1	US-08-237-919-1
3	14	1.7	2808	4	US-08-732-429-1
4	14	1.7	2808	4	US-09-798-267-1
5	14	1.7	2808	5	PCT-US95-05518-1
6	13	1.6	3048	1	US-08-188-228-47
7	13	1.6	3048	1	US-08-332-643-41
8	13	1.6	3048	1	US-08-332-638-47
9	12	1.4	436	1	US-08-397-787-318
10	11	1.3	117	1	US-08-188-228-9
11	11	1.3	117	1	US-08-332-643-9
12	11	1.3	117	1	US-08-332-638-9

13	11	1.3	938	4	US-09-976-594-1052	Sequence 1052, Ap
14	11	1.3	1369	1	US-08-188-228-51	Sequence 51, Appl
15	11	1.3	1369	1	US-08-332-643-45	Sequence 45, Appl
16	11	1.3	1369	1	US-08-332-638-51	Sequence 51, Appl
17	10	1.2	326	4	US-09-389-681-288	Sequence 288, App
18	10	1.2	326	4	US-09-620-405B-288	Sequence 288, App
19	10	1.2	326	4	US-09-339-338-288	Sequence 288, App
20	10	1.2	326	4	US-09-433-826B-288	Sequence 288, App
21	10	1.2	326	4	US-09-604-287A-288	Sequence 288, App
22	10	1.2	326	4	US-09-834-759-288	Sequence 288, App
23	9	1.1	402	1	US-08-346-611-1	Sequence 1, Appli
24	9	1.1	402	1	US-08-346-611-3	Sequence 3, Appli
25	9	1.1	402	1	US-08-794-494-1	Sequence 1, Appli
26	9	1.1	402	2	US-08-794-494-3	Sequence 3, Appli
27	9	1.1	813	4	US-09-641-638-100	Sequence 100, App
28	9	1.1	943	4	US-09-641-638-99	Sequence 99, Appl
29	9	1.1	1484	2	US-08-037-816A-17	Sequence 17, Appl
30	9	1.1	1484	2	US-08-037-816A-21	Sequence 21, Appl
31	9	1.1	1484	2	US-08-530-146-17	Sequence 17, Appl
32	9	1.1	1484	2	US-08-530-146-21	Sequence 21, Appl
33	9	1.1	1571	2	US-08-037-816A-25	Sequence 25, Appl
34	9	1.1	1571	2	US-08-530-146-25	Sequence 25, Appl
35	9	1.1	1818	4	US-09-475-515-59	Sequence 59, Appl
36	9	1.1	1818	4	US-09-475-515-61	Sequence 61, Appl
37	9	1.1	1818	4	US-09-475-515-62	Sequence 62, Appl
38	9	1.1	1836	4	US-09-475-515-41	Sequence 41, Appl
39	9	1.1	1836	4	US-09-475-515-44	Sequence 44, Appl
40	9	1.1	1836	4	US-09-475-515-47	Sequence 47, Appl
41	9	1.1	1863	4	US-09-475-515-63	Sequence 63, Appl
42	9	1.1	1944	4	US-09-475-515-37	Sequence 37, Appl
43	9	1.1	1944	4	US-09-475-515-38	Sequence 38, Appl
44	9	1.1	1944	4	US-09-475-515-40	Sequence 40, Appl
45	9	1.1	1944	4	US-09-475-515-43	Sequence 43, Appl

ALIGNMENTS

RESULT 1
US-09-016-434-1348
; Sequence 1348, Application US/09016434
; Patent No. 6500938
; GENERAL INFORMATION:
; APPLICANT: Janice Au-Young
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION
; NUMBER OF SEQUENCES: 1490
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,434
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: FA-0002 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555

TELEFAX: (650) 845-4166
 INFORMATION FOR SEQ ID NO: 1348:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3171 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 LIBRARY: GENBANK
 CLONE: G35322
 US-09-016-434-1348

Alignment Scores:

Pred. No.: 0 Length: 3171
 Score: 829.00 Matches: 829
 Percent Similarity: 100.00% Conservativeness: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 4 Gaps: 0

US-09-916-849A-1 (1-829) x US-09-016-434-1348 (1-3171)

QY 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuLeuValCysTrpLeuGln 20
 DB 54 ATGGGGCTCCCTCGTGAGACTCTCGGCTCTCTCCCTCTCCAGGTTTGTGGCTGCAG 113
 QY 21 CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
 DB 114 TGGCGGCTCCGAGCGGTGCGGGCGGTCTTCAGGGAGGCTGAAGTGACCTTGGAGCG 173
 QY 41 GlyGlyAlaGluGlnGluProGlyGlnAlaLeuGlyValPheMetGlyCysProGly 60
 DB 174 GAGGCGCGGAGAGAGCCGCGCAGGCGCTGGGGAAGTATTCATGGCTGCCCTGGG 233
 QY 61 GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr 80
 DB 234 CAAGAGCCAGCTCTGTTAGCACTGATAATGATGACTTCACTGTGCGGAATGGCGAGACA 293
 QY 81 ValGlnGluArgArgSerLeuGlyGluArgAsnProLeuLeuValPheProSerLysArg 100
 DB 294 GTCCAGGAAGAGGTCATGAGGAAGAGATTCATTGAAGATCTTCCCATCCAAACGT 353
 QY 101 IleLeuArgArgHisLysArgAspTrpValAlaProIleSerValProGluAsnGly 120
 DB 354 ATCTTACGAGACACAGAGAGATTGGGTGGTGTCTCCAAATATCTGCTCGAATGGC 413
 QY 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
 DB 414 AAGGGTCCCTTCCCCAGAGACTGAATCAGCTCAAGTCTTAATAAAGATAGACACCAAG 473
 QY 141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProGluGlyValPheAlaVal 160
 DB 474 ATTTCTACAGCATACGGGGCGGGGGCAGACAGCCCCCTTGAGGGTGTCTTCCTGTA 533
 QY 161 GluLysGluThrGlyTrpLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys 180
 DB 534 GAGAGGACACAGGCTGGTGTGTTGTAATAAGCACTGAGCCGGGAGAGATTGCCAAG 593
 QY 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
 DB 594 TATGAGCTCTTGGCCACCCTGTGTGAGAGATGTGTGCTCAGTGGAGAGCCCCATGAAC 653
 QY 201 IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
 DB 654 ATCTCCATCATGTGACCCAGCAGAGATGACCAAGCCCAAGTTTACCCAGGACACCTTC 713
 QY 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
 DB 714 CGAGGGAGGTGCTTACAGGGATCTTACCAGGTACTTCTGTGATGAGGTGACAGCCACA 773
 QY 241 AspGluAspAlaIleTyrThrValAsnGlyValAlaIleTyrSerIleHisSerGln 260
 DB 774 GATGAGGATGATGCCATCTACACCTACAAATGGGTGGTGTGCTTACTTCCATCCATAGCAA 833

QY 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
 DB 834 GAACCAAGGAGCCACACACACCTCATGTTTCAATTCACGGAGCAGACGACCATCAGC 893
 QY 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
 DB 894 GTCATCTCCAGTGGCTGGACCGGAAAGTCCCTGAGTGACACTGACCATCCAGGCC 953
 QY 301 ThrAspMetAspGlyAspGlySerThrThrAlaValAlaValValGluLeuLeuAsp 320
 DB 954 ACAGACATGATGGGACCGCTCCACCAACACCGCAGTGGCAGTAGTGAGATCCTTGAT 1013
 QY 321 AlaAsnAspAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
 DB 1014 GCCAATGACAATGCTCCATGTTTACCCCGGAGAGTACGAGCCCATGTGCTTGAGAT 1073
 QY 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
 DB 1074 GCAGTGGGCCATGAGGTGCAGAGGCTGACGTCACGTGACCTGGAGCCGCCCACTCACA 1133
 QY 361 AlaTrpArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
 DB 1134 GCGTGGCGTCCCACTACCTTATCATGGCGGTGACGACGGGACCATTTTACCATCACC 1193
 QY 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
 DB 1194 ACCACCTCCAGAGCACCAGGCGCATCTGACACCGAGAGGTTTGGATTTTGGAGCC 1253
 QY 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
 DB 1254 AAAACCCAGCACACCTGTAGCTTGAAGTGACCAACGAGGCCCTTTTGTGCTGAAGCTC 1313
 QY 421 ProThrSerThrAlaThrIleValHisValGluAspValAsnGluAlaProValPhe 440
 DB 1314 CCAACCTCCACACCCACCATAGTGGTCCACGTGGAGGATGTGAATGAGGACCTGTGTGT 1373
 QY 441 ValProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
 DB 1374 GTCCACCTCCAAAGTCGTTGAGGTCCAGGAGGCAATCCCACTGGGAGGCTGTGTGT 1433
 QY 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
 DB 1434 GTCTACACTCAGAGACCCCTGACAAAGAGAGATCAAAAGATCAGCTACCGCATCTCGAGA 1493
 QY 481 AspProAlaGlyTrpLeuAlaValMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
 DB 1494 GACCCAGCAGGTGGCTAGCCATGGACCCAGACAGTGGGAGGTACAGCTGTGGGCACC 1553
 QY 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
 DB 1554 CTCGACCGTGAGGATGAGCAGTGTGTGAGGAACAACATCTATGAAGTCATGTGCTTGGCC 1613
 QY 521 MetAspAsnGlySerProThrThrThrGlyThrGlyThrLeuLeuLeuLeuLeuLeu 540
 DB 1614 ATGCAATGAGAGAGCCCTCCCACTGGCAGCGGAAACCTTCTGCTAACACTGATTGAT 1673
 QY 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
 DB 1674 GTCAACGACCATGGCCAGTCCCTGAGCCCGCTCAGATCACCATCTGCAACCAAGCCCT 1733
 QY 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
 DB 1734 GTGGCCACGTGTGTGAACATCAGGACAAAGGACCTGTCTCCCAACACCTTCCCTTCCAG 1793
 QY 581 AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGluGlyAsp 600
 DB 1794 GCCAGCTCAGATGACTCAGACATCTACTGACCGGAGAGGTCAACGAGAGGTGAC 1853
 QY 601 ThrValValLeuSerLeuLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
 DB 1854 ACAGTGTGCTTGTCCCTGAAGAAGTCTCTGAAGAGGATACATATGACGTGACCTTTCT 1913

QY 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
DB 1914 CTGTCTGACCATGGCAACAAGAGCAGCTGACGGTGATCAGGGCCACTGTGTGGACTGC 1973
QY 641 HisGlyHisValGluThrCysProGlyProThrLysGlyGlyPheIleLeuProValLeu 660
DB 1974 CATGCCCATGTGCAACCTGCCCTGAGCCCTGGAAGAGGTTTCATCTCCCTGTGTG 2033
QY 661 GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuValLeuValArgLysLys 680
DB 2034 GGGGCTGTCTGGCTCTGCTGTCTCTCTGTGTGTGTCTTTTGTGTGGAAGAAG 2093
QY 681 ArgLysIleLysGluProLeuLeuLeuProGluAspThrArgAsnValPheTyr 700
DB 2094 CGGAAGATCAAGAGGCCCTCTACTCCAGAGATGACCCCGTGACAACGCTCTTCTAC 2153
QY 701 TyrGlyGluGluGlyGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
DB 2154 TATGGGAAGAGGGGGTGGCGAGAGGACGAGGACTATGATCACCACCTCCACCGA 2213
QY 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
DB 2214 GGTCTGGAGGCCAGCGCGAGGTGTCTCCGCAATGACGTGGCACCACCAACCATCCTCCG 2273
QY 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
DB 2274 ACACCATGTACCGTCTAGGCGACCAACCCAGATGAATCGCACTTTATTAATTGAG 2333
QY 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
DB 2334 AACCTGAAGCGGTAAACACAGACCCACAGCCCGCCCTACGACACCTCTTGTGTGTC 2393
QY 781 AspTyrGluGlySerGlySerAspAlaSerLeuSerSerLeuThrSerSerAlaSer 800
DB 2394 GACTATGAGGAGCGGGTCCGACCGCGTCCCTGAGCTCCCTCACCCTCTCCGCCCTCC 2453
QY 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTyrGlySerArgPheLysLysLeuAla 820
DB 2454 GACCAGACCAAGATACGATTATCTGACGAGTGGGCGAGCGGCTTCAAGAGCTGCA 2513
QY 821 AspMetTyrGlyGlyGlyGluAspAsp 829
DB 2514 GACATGTACGTGGTGGGGGAGGAGCAG 2540
RESULT 2
US-08-237-919-1
; Sequence 1, Application US/08237919
; Patent No. 5610281
; GENERAL INFORMATION:
; APPLICANT: Brenner, Michael B
; TITLE OF INVENTION: Methods and Compositions for
; Modulating Heterotypic E-cadherin Interactions with T Lymphocytes
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R
; REGISTRATION NUMBER: 36,637

REFERENCE/DOCKET NUMBER: B0801/7023
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2808 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: liver
FEATURE:
NAME/KEY: CDS
LOCATION: 109..2745
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 109..558
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 559..2742
FEATURE:
NAME/KEY: cytoplasmic tail
LOCATION: 2296..2746
FEATURE:
NAME/KEY: HAV tripeptide
LOCATION: 800..808
FEATURE:
NAME/KEY: transmembrane
LOCATION: 2225..2295
US-08-237-919-1
Alignment Scores:
Pred. No.: 0.0011 Length: 2808
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.69% Indels: 0
DB: 1 Gaps: 0
US-09-916-849A-1 (1-829) x US-08-237-919-1 (1-2808)
QY 815 ArgPheLysLysLeuAlaAspMetTyrGlyGlyGluAsp 828
DB 2698 CGCTTCAAGAGCTGGCTGACATGTACGGAGCGGCGAGGAC 2739
RESULT 3
US-08-732-429-1
; Sequence 1, Application US/08732429
; Patent No. 6300080
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic
; E-cadherin Interactions with T Lymphocytes
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/732,429

```

; FILING DATE: herewith
; PRIOR APPLICATION NUMBER: 08/237,919
; APPLICATION DATE: 3 May 1994
; FILING DATE: 3 May 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7023
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2808 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: liver
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 109..2745
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 109..558
; NAME/KEY: mat_peptide
; LOCATION: 559..2742
; FEATURE:
; NAME/KEY: cytoplasmic tail
; LOCATION: 2296..2746
; NAME/KEY: HAV tripeptide
; LOCATION: 800..808
; FEATURE:
; NAME/KEY: transmembrane
; LOCATION: 2225..2295
; US-08-732-429-1
;
Alignment Scores:
Pred. No.: 0.0011 Length: 2808
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.69% Indels: 0
DB: 4 Gaps: 0
;
US-09-916-849A-1 (1-829) x US-08-732-429-1 (1-2808)
QY 815 ArgpHelvSLysLeuAlaAspMetTVrGlyGlyGluasp 828
Db 2698 CGCTTCAAGAGCTGGCTGACATGTACGGAGGCGCGAGGAC 2739
;
RESULT 4
US-09-798-267-1
; Sequence 1, Application US/09798267
; Patent No. 6406870
; GENERAL INFORMATION:
; APPLICANT: Brenner, Michael
; APPLICANT: Ceppek, Karyn
; TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic E-cadherin
; FILE REFERENCE: L0560/7008EP
; CURRENT APPLICATION NUMBER: US/09/798,267
; CURRENT FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: US 08/237,919
; PRIOR FILING DATE: 1994-05-03
; PRIOR APPLICATION NUMBER: PCT/US 95/05518
; PRIOR FILING DATE: 1995-05-03

```

```

; PRIOR APPLICATION NUMBER: US 08/732,429
; PRIOR FILING DATE: 1996-11-01
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2808
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (109)..(2742)
; NAME/KEY: mat_peptide
; LOCATION: (2740)..()
; NAME/KEY: misc feature
; LOCATION: (800)..(808)
; OTHER INFORMATION: HAV tripeptide
; NAME/KEY: misc feature
; LOCATION: (2225)..(2295)
; OTHER INFORMATION: transmembrane
; NAME/KEY: misc structure
; LOCATION: (2296)..(2746)
; OTHER INFORMATION: cytoplasmic tail
; NAME/KEY: sig_peptide
; LOCATION: (109)..(558)
; US-09-798-267-1
;
Alignment Scores:
Pred. No.: 0.0011 Length: 2808
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.69% Indels: 0
DB: 4 Gaps: 0
;
US-09-916-849A-1 (1-829) x US-09-798-267-1 (1-2808)
QY 815 ArgpHelvSLysLeuAlaAspMetTVrGlyGlyGluasp 828
Db 2698 CGCTTCAAGAGCTGGCTGACATGTACGGAGGCGCGAGGAC 2739
;
RESULT 5
PCT-US95-05518-1
; Sequence 1, Application PC/TUS9505518
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic
; E-cadherin Interactions with T Lymphocytes
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/05518
; FILING DATE: herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/237,919
; FILING DATE: 3 May 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7023
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441

```



```

/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 3048 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
US-08-332-643-41
Alignment Scores:
Pred. No.: 0.0113 Length: 3048
Score: 13.00 Matches: 13
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.57% Indels: 0
DB: 1 Gaps: 0
US-09-916-849A-1 (1-829) x US-08-332-643-41 (1-3048)
RESULT 8
QY 815 ArgPheLysLeuAlaAspMetTyrGlyGlyGlu 827
DB 2718 AGATTCAAGAGCTGCGGACATGTATGGAGTGTGAA 2756
SEQUENCE CHARACTERISTICS:
LENGTH: 3048 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-332-638-47
Alignment Scores:
Pred. No.: 0.0162 Length: 436
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.45% Indels: 0
DB: 4 Gaps: 0
US-09-916-849A-1 (1-829) x US-09-397-787-318 (1-436)
RESULT 9
QY 815 ArgPheLysLeuAlaAspMetTyrGlyGlyGly 826
DB 290 CGTTCAAGAACTTGTGACATGTATGGTGGAGT 325
SEQUENCE CHARACTERISTICS:
LENGTH: 436
TYPE: DNA
ORGANISM: Homo sapien
US-09-397-787-318
Alignment Scores:
Pred. No.: 0.0162 Length: 436
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.45% Indels: 0
DB: 4 Gaps: 0
US-09-916-849A-1 (1-829) x US-09-397-787-318 (1-436)
RESULT 10
QY 815 ArgPheLysLeuAlaAspMetTyrGlyGlyGly 826
DB 290 CGTTCAAGAACTTGTGACATGTATGGTGGAGT 325
SEQUENCE CHARACTERISTICS:
LENGTH: 436
TYPE: DNA
ORGANISM: Homo sapien
US-09-397-787-318
Alignment Scores:
Pred. No.: 0.0162 Length: 436
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.45% Indels: 0
DB: 4 Gaps: 0
US-09-916-849A-1 (1-829) x US-09-397-787-318 (1-436)

```

```

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.57% Indels: 0
DB: 1 Gaps: 0
US-09-916-849A-1 (1-829) x US-08-332-638-47 (1-3048)
QY 815 ArgPheLysLeuAlaAspMetTyrGlyGlyGlu 827
DB 2718 AGATTCAAGAGCTGCGGACATGTATGGAGTGTGAA 2756
RESULT 9
US-09-397-787-318
; Sequence 318, Application US/093977787
; Patent No. 6468758
; GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR OVARIAN
; FILE REFERENCE: 210121.466C2
; CURRENT FILING DATE: 1999-09-16
; CURRENT APPLICATION NUMBER: US/09/397,787
; NUMBER OF SEQ ID NOS: 334
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 318
; LENGTH: 436
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-397-787-318
Alignment Scores:
Pred. No.: 0.0162 Length: 436
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.45% Indels: 0
DB: 4 Gaps: 0
US-09-916-849A-1 (1-829) x US-09-397-787-318 (1-436)
QY 815 ArgPheLysLeuAlaAspMetTyrGlyGlyGly 826
DB 290 CGTTCAAGAACTTGTGACATGTATGGTGGAGT 325
RESULT 10
US-08-188-228-9
; Sequence 9, Application US/08188228
; Patent No. 5597725
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; STREET: 6300 Sears Tower, 233 S. Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/049,460
; FILING DATE: 01-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,643
; FILING DATE: 17 APR 1992
; APPLICATION NUMBER: US/08/049,460
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5646250and, Greta E.
; REGISTRATION NUMBER: 35,502
; REFERENCE/DOCKET NUMBER: 31340
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 47:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3048 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-332-638-47
Alignment Scores:
Pred. No.: 0.0113 Length: 3048
Score: 13.00 Matches: 13
Percent Similarity: 100.00% Conservative: 0

```

```

; FILING DATE: 19 APR 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,643
; FILING DATE: 17 APR 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: NO. 5597725and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31340
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-188-228-9

Alignment Scores:
Pred. No.: 0.0427 Length: 117
Score: 11.00 Matches: 11
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 1.33%
Indels: 0
DB: 1
Gaps: 0

US-09-916-849A-1 (1-829) x US-08-188-228-9 (1-117)

QY 777 LeuLeuValPheAspTyrCluGlySerGlySer 787
Db 4 CTGCTGGTCTTCGACTACGAGGCGGTTCT 36

RESULT 11
US-08-332-643-9
; Sequence 9, Application US/08332643
; Patent No. 5639634
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Bicknell
; STREET: Two First National Plaza, 20 South Clark
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/332,643
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/872,643
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: NO. 5639634and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/30795
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 346-5750
; TELEFAX: (312) 984-9740
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 9:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-332-643-9

Alignment Scores:
Pred. No.: 0.0427 Length: 117
Score: 11.00 Matches: 11
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 1.33%
Indels: 0
DB: 1
Gaps: 0

US-09-916-849A-1 (1-829) x US-08-332-643-9 (1-117)

QY 777 LeuLeuValPheAspTyrCluGlySerGlySer 787
Db 4 CTGCTGGTCTTCGACTACGAGGCGGTTCT 36

RESULT 12
US-08-332-638-9
; Sequence 9, Application US/08332638
; Patent No. 5646250
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Borun
; STREET: 6300 Sears Tower, 233 S. Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/332,638
; FILING DATE: 01-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,643
; FILING DATE: 17 APR 1992
; APPLICATION NUMBER: US/08/049,460
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: NO. 5646250and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31340
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-332-638-9

Alignment Scores:
Pred. No.: 0.0427 Length: 117
Score: 11.00 Matches: 11
Percent Similarity: 100.00%

```

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.33% Indels: 0
DB: 1 Gaps: 0

US-09-916-849A-1 (1-829) x US-08-332-638-9 (1-117)

QY 777 LeuLeuValPheAspTyrGluGlySerGlySer 787
|||||
DB 4 CTGCTGCTTCGACTACGAGCGCGGTCT 36

RESULT 13

US-09-976-594-1052
; Sequence 1052, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 1052
; LENGTH: 938
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 272273.1
US-09-976-594-1052

Alignment Scores: Pred. No.: 0.321 Length: 938
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.33% Indels: 0
DB: 4 Gaps: 0

US-09-916-849A-1 (1-829) x US-09-976-594-1052 (1-938)

QY 815 ArgPheLysLysLeuAlaAspMetTyrGlyGly 825
|||||
DB 349 CGTTTCAAAAAGCTTCCGATATGATGGGGT 381

RESULT 14

US-08-188-228-51
; Sequence 51, Application US/08188228
; Patent No. 5597725
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Borun
; STREET: 6300 Sears Tower, 233 S. Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/188,228
; FILING DATE:
; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/049,460
; FILING DATE: 19 APR 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,643
; FILING DATE: 17 APR 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5597725and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31340
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1369 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
US-08-188-228-51
Alignment Scores: Pred. No.: 0.462 Length: 1369
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.33% Indels: 0
DB: 1 Gaps: 0

US-09-916-849A-1 (1-829) x US-08-188-228-51 (1-1369)

QY 815 ArgPheLysLysLeuAlaAspMetTyrGlyGly 825
|||||
DB 1190 CGATTCAAAAAGCTTGCAGATATGATGGAGGA 1222

RESULT 15

US-08-332-643-45
; Sequence 45, Application US/08332643
; Patent No. 5639634
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Ricknell
; STREET: Two First National Plaza, 20 South Clark
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/332,643
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/872,643
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5639634and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/30795
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 346-5750
; TELEFAX: (312) 984-9740


```
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 45:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1369 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
US-08-332-643-45

Alignment Scores:
Pred. No.:      0.462      Length:      1369
Score:          11.00      Matches:      11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:      1.33%   Indels:      0
DB:                1      Gaps:        0

US-09-916-849A-1 (1-829) x US-08-332-643-45 (1-1369)
QY      815 ArgPheLysLysLeuAlaAspMetTyrGlyGly 825
      |||||
Db      1190 CGATTCARAAAGCTTCAGATATGATGGAGGA 1222
```

Search completed: September 23, 2004, 07:41:11
Job time : 176 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 23, 2004, 06:21:13 ; Search time 1807 Seconds
(without alignments)
2323.654 Million cell updates/sec

Title: US-09-916-849A-1

Perfect score: 829

Sequence: 1 MGLPRGLASILLLOVCLQ.....NEWGRFKKLADMYGGEDD 829

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3337386 seqs, 2532474682 residues

Word size: 1

Total number of hits satisfying chosen parameters: 6665927

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool.h/US09916849/runat_21092004_145838_7417/app.query.fasta_1.967
-DS=Published Applications NA -QMT=fastcap -SUFFIX=oligo.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOGPXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=quality -THR_MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFM=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09916849 @CGN 1.1.593 @runat_21092004_145838_7417
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/prodata/2/pubpna/PTC_NEW_PUB.seq:
3: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/prodata/2/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/prodata/2/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/prodata/2/pubpna/PTC_PUBCOMB.seq:
7: /cgn2_6/prodata/2/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/prodata/2/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/prodata/2/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/prodata/2/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq:
14: /cgn2_6/prodata/2/pubpna/US10A_PUBCOMB.seq:
15: /cgn2_6/prodata/2/pubpna/US10B_PUBCOMB.seq:
16: /cgn2_6/prodata/2/pubpna/US10C_PUBCOMB.seq:
17: /cgn2_6/prodata/2/pubpna/US10_NEW_PUB.seq:
18: /cgn2_6/prodata/2/pubpna/US60_NEW_PUB.seq:
19: /cgn2_6/prodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
<p>* Query Sequence</p>					

1	829	100.0	3170	10	US-09-975-723A-2	Sequence 2, Appli
2	829	100.0	3171	9	US-09-905-983-49	Sequence 49, Appl
3	829	100.0	3171	9	US-09-880-107-3759	Sequence 3759, Ap
4	829	100.0	3171	13	US-10-058-270A-125	Sequence 125, App
5	829	100.0	3171	13	US-10-342-887-669	Sequence 669, App
6	829	100.0	3171	13	US-10-172-118-669	Sequence 669, App
7	829	100.0	3171	15	US-10-174-677-115	Sequence 115, App
8	829	100.0	3171	15	US-10-158-123-2	Sequence 2, Appli
9	829	100.0	3171	16	US-10-295-027-782	Sequence 782, App
10	829	100.0	3171	16	US-10-295-027-799	Sequence 799, App
11	829	100.0	3171	16	US-10-295-027-862	Sequence 862, App
12	829	100.0	3171	16	US-10-295-027-895	Sequence 895, App
13	829	100.0	3171	16	US-10-305-720-1348	Sequence 1348, Ap
14	829	100.0	3171	17	US-10-775-920-188	Sequence 188, App
15	829	100.0	3777	13	US-10-116-802-6	Sequence 6, Appli
16	728	87.8	3205	13	US-10-229-345-7	Sequence 7, Appli
17	728	87.8	3205	13	US-10-274-177-7	Sequence 7, Appli
18	728	87.8	3205	15	US-10-301-822-17	Sequence 17, Appl
19	728	87.8	3205	16	US-10-295-027-337	Sequence 337, App
20	728	87.8	3205	16	US-10-295-027-1029	Sequence 1029, Ap
21	728	87.8	3205	16	US-10-087-080-13	Sequence 13, Appl
22	728	87.8	3205	17	US-10-650-112-7	Sequence 2, Appli
23	728	87.8	3205	17	US-10-775-920-186	Sequence 186, App
24	715	86.2	3219	9	US-09-895-652-17	Sequence 17, Appl
25	659	79.5	3632	17	US-10-775-920-187	Sequence 187, App
26	102	12.3	306	10	US-09-930-213-331	Sequence 331, App
27	89	10.7	3314	16	US-10-062-674-1863	Sequence 1863, Ap
28	80	9.7	247	15	US-10-029-386-15020	Sequence 15020, A
29	80	9.7	555	15	US-10-029-386-1318	Sequence 1318, Ap
30	69	8.3	210	15	US-10-029-386-17424	Sequence 17424, A
31	69	8.3	574	15	US-10-029-386-3724	Sequence 3724, Ap
32	54	6.5	255	10	US-09-930-213-381	Sequence 381, App
33	52	6.3	160	15	US-10-158-123-3	Sequence 3, Appli
34	47	5.7	406	16	US-10-027-632-571	Sequence 571, App
35	30	3.6	767	13	US-10-027-632-16189	Sequence 16189, A
36	30	3.6	767	16	US-10-027-632-16189	Sequence 16189, A
37	21	2.5	522	9	US-09-920-455-77	Sequence 77, Appl
38	19	2.3	60	10	US-09-908-975-13599	Sequence 13599, A
39	15	1.8	195	15	US-10-029-386-16396	Sequence 16396, A
40	15	1.8	542	15	US-10-029-386-2696	Sequence 2696, Ap
41	15	1.8	1351	15	US-10-198-846-12909	Sequence 12909, A
42	15	1.8	2768	9	US-09-905-983-4	Sequence 4, Appli
43	15	1.8	2768	9	US-09-905-983-6	Sequence 6, Appli
44	15	1.8	4396	12	US-10-152-319A-1300	Sequence 1300, Ap
45	15	1.8	4778	13	US-10-342-887-1027	Sequence 1027, Ap

ALIGNMENTS

RESULT 1
US-09-975-723A-2
; Sequence 2, Application US/09975723A
; Publication No. US20030108529A1
; GENERAL INFORMATION:
; APPLICANT: Nackman, Gary
; APPLICANT: Foly, Ramsey
; TITLE OF INVENTION: Improvement of Endothelial Cell-Cell
; TITLE OF INVENTION: Conesion
; FILE REFERENCE: 601-1-101N
; CURRENT APPLICATION NUMBER: US/09/975,723A
; PRIOR FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: 60/241,216
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/243,693
; PRIOR FILING DATE: 2000-10-27
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 3170
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-975-723A-2

Alignment Scores:			
Pred. No.:	0	Length:	3170
Score:	829.00	Matches:	829
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	10	Gaps:	0
US-09-916-849A-1 (1-829) x US-09-975-723A-2 (1-3170)			
Qy	1	MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuLeuGlnValCysTrpLeuGln	20
Db	54	ATGGGGCTCCCTCGTGGACCTCTCGCGTCTCTCTCTCTCTCCAGGTTCGTGGTGCAG	113
Qy	21	CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla	40
Db	114	TGGCGGGCTCCGAGCGGTGCGGGCGGTCTTCAGGGAGGCTGAAGTACCTTTGGAGGGC	173
Qy	41	GlyGlyAlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly	60
Db	174	GGAGGCGCGAGCAGGAGCCCGGCAGCGCTGGGAAAGTATTCATGGGCTCCCGCTGG	233
Qy	61	GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr	80
Db	234	CAAGAGCCAGCTCTGTTAGCACTGATTAATGATGACTTCATGTGGGAATGGCGAGCA	293
Qy	81	ValGlnGluArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg	100
Db	294	GTCCAGGAAGAAGGTCACTGAAGAAAGGAATCCATTGAAGATCTTCCCATCCAAAGT	353
Qy	101	IleLeuArgHisLysArgAspTrpValValAlaProIleSerValProGluAsnGly	120
Db	354	ATCTTACGAAGACACAGAGAGATTGGGTGGTGTCTCCAAATATCTGTCCCTGAAATGGC	413
Qy	121	LysGlyProPheProGluArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys	140
Db	414	AAAGGTCCCTTCCCCAGAGACTGAATCAGCTCAAGCTTAATAAGATAGAGACACCAAG	473
Qy	141	IlePheTy-Ser-IleThrGlyProGlyAlaAspSerProGluGlyValPheAlaVal	160
Db	474	ATTTCCTACAGCATCAGCGGGCGGGGCGACAGCCCCCTCAGGGTGTCTTCGTGTGA	533
Qy	161	GluLysGluThrGlyTrpLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys	180
Db	534	GAGAAGGACACAGGCTGGTGTGTGTGAATGAAGCCACTGGACCGGGGAGGATTGGCAAG	593
Qy	181	TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn	200
Db	594	TATGAGCTCTTTGGCCAGCGTGTGTACAGAGAAATGGTCCTCAGTGGAGGACCCCATGAAC	653
Qy	201	IleSer-IleLeuValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe	220
Db	654	ATTCCTCATGTGTACCGACCGAATGACCAAGGCCAAGTTTACCCAGGACACCTTC	713
Qy	221	ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr	240
Db	714	CGAGGGAGTGTCTTAGGGGAGTCTCACAGGTACTTCTGTGATGAGGTGACAGCCACA	773
Qy	241	AspGluAspAlaIleTyrThrTyAsnGlyValValAlaTySerIleHisSerGln	260
Db	774	GATGAGGATGATGCCATCTACACCTACAAATGGGGTGGTGTCTTACTTCATCCATAGCCAA	833
Qy	261	GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer	280
Db	834	GAACCAAGGACCCACAGACCTCATGTCTCAATTCACAATTCACCGGAGCACAGGACCATCAGC	893
Qy	281	ValIleSerSerGlyLeuAspArgGluLysValProGluTyThrThrLeuThrIleGlnAla	300
Db	894	GTCAATCTCAGTGGCTGGACCGGGAAAAGTCCCTGAGTACACATGACATCTCCAGGCC	953
Qy	301	ThrAspMetAspGlyAspGlySerThrThrThrAlaValAlaValGluIleLeuAsp	320

Qy	681	ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr	700
Db	2094	CGGAAGATCAGGAGCCCTCTACTCCAGAAGATGACCCGTGCAACGCTCTTCTAC	2153
Qy	701	TyrGlyGluGluGlyGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisArg	720
Db	2154	TATGGCGAAGGGGGGTGGCAGAGAGGACGAGACTATGACATCACCAGCTCCACCGA	2213
Qy	721	GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro	740
Db	2214	GGTCTGAGGCCAGCGCGGAGGTGGTCTCCGCAATGACGTGGCACCACCAATCATCCCG	2273
Qy	741	ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu	760
Db	2274	ACACCCATGTACCGTCTCTAGSCCAGCCACCAGATGAATCGSCAACTTTATAATTGAG	2333
Qy	761	AsnLeuLysAlaAlaAsnThrAspProThrAlaProCotYrAspThrLeuLeuValPhe	780
Db	2334	AACCTGAAGGGCGGTATACACAGACCCACAGCCCCGCCCTACGACACCCCTCTTGGTGTC	2393
Qy	781	AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerLeuThrSerSerAlaSer	800
Db	2394	GACTATGAGGGCAGCGGCTCCGACGCGCGTCCCTGAGCTCCCTCACTCTCTCGCGCTCC	2453
Qy	801	AspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGlySerArgPheLysLysLeuAla	820
Db	2454	GACCAAGACCAAGATTACGATTATCTGAACGAGTGGGGCAGCCGCTTCAAGAAAGCTGCA	2513
Qy	821	AspMetTyrGlyGlyGluAspAsp	829
Db	2514	GACATGTACGTGGCGGGAGGACAC	2540

RESULT 2

```

US-09-905-983-49
; Sequence 49, Application US/09905983
; Patent No. US20020045591A1
; GENERAL INFORMATION:
; APPLICANT: Geiger, Benjamin
; APPLICANT: Ben-Ze'ev, Avri
; APPLICANT: Sadot, Einat
; TITLE OF INVENTION: METHODS AND THERAPEUTIC COMPOSITIONS FOR TREATING CANCER
; FILE REFERENCE: 01/22326
; CURRENT APPLICATION NUMBER: US/09/905,983
; CURRENT FILING DATE: 2001-09-28
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 49
; LENGTH: 3171
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-905-983-49

```

Alignment Scores:			
Pred. No.:	0	Length:	3171
Score:	829.00	Matches:	829
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0

US-09-916-849A-1 (1-829) x US-09-905-983-49 (1-3171)

Qy	1	MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuLeuGlnValCysTyrLeuGln	20
Db	54	ATGGGCTCCCTCGTGGACCTCTCCGCTCTCTCTCTCCAGGTTGCTGGTCGAG	113
Qy	21	CysAlaAlaSerClyProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla	40
Db	114	TGGCGGCTCCAGCGCGTGGCGCGCTTCCTAGGAGGCTGAAGTGACCTTGGAGGCG	173
Qy	41	GlyGlyAlaGluGlnGlyProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly	60

Qy 421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
Db 1314 CCAACCTCCACAGCACCATAGTGGTCCACGTGGAGGATGTGAATGAGGACCATGTGTGTTT 1373
Qy 441 ValProSerLysValValGluValGluGluIleProThrGlyGluProValCys 460
Db 1374 GTCCACCCCTCCAAAGTCGTTGAGTCCAGAGGCAATCCCACTGGGAGCCGTGTGT 1433
Qy 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db 1434 GTCTACTGCAGAGAGCCCTGCACAGAGAAATCAAAAGATCAGTACCGCATCTCTGAGA 1493
Qy 481 AspProHlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db 1494 GACCCAGCGGGTGGCTAGCCATGGACCCAGACAGTGGGAGGTCAAGTGTGGGCACC 1553
Qy 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db 1554 CTCGACCGTGGAGTGAAGCTTTGTGAGGAACAACATCTATGAAGTCATGTGTCTGGCC 1613
Qy 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAsp 540
Db 1614 ATGGACAATGGAGCCCTCCACACCTGGCACGGGAACCTTCTGTCTAACTGATGAT 1673
Qy 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 1674 GTCAACACCATGGCCAGTCCCTGAGCCCGTCAGATCACCATCTGCAACCAAGCCCT 1733
Qy 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
Db 1734 GTGGCCACCTGTGAACATCAGGACAGGACCTGTCTCCACACCTCCCTTTCCAG 1793
Qy 581 AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGlyAsp 600
Db 1794 GCCCAGTCTCAGATGACTCAGACATCTCTGACCGCAGAGGTCAACGAGGAGGTGAC 1853
Qy 601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
Db 1854 ACAGTGTCTGTCCCTGAGAGATTCCTGAAGCAGGATACATATGAGTGCACCTTTCT 1913
Qy 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
Db 1914 CTGTCTACCATGGCAACAAGAGCAGCTGACCGTGTATCAGGGCCACTGTGTGCGACTGC 1973
Qy 641 HisGlyHisValGluThrCysProGlyProTyrPheIleLeuProValLeu 660
Db 1974 CATGGCCATGTGNAACCTGCTCCTCTGCTGTGTCTGTTTGTGTGTGAGAAAGAAG 2093
Qy 661 GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysLys 680
Db 2034 GGGGCTGTCTGTGCTGTCT 2093
Qy 681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
Db 2094 CGGAAGATCAAGAGGCCCT 2153
Qy 701 TyrGlyGluGlyGlyGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
Db 2154 TATGGCAAGAGGGGGTGGCGAAGAGACGACGACTATGACATCACCCAGCTCCACCGA 2213
Qy 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleLeuPro 740
Db 2214 GGTCTGAGGCGCAGCGCGAGGTGGTCTCCGCAATGACGTGGCACCACCATCATCCCG 2273
Qy 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleLeuGlu 760
Db 2274 ACACCCATGTACCGTCTAGGCCAGCCACCCAGATGAAATCGCAACTTTATAATGAG 2333
Qy 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
Db 2334 AACCTGAAGCGGCTAAACACAGACCCCCACAGCCCCCCTTACGACACCCCTCTTGGGTTC 2393

Qy 781 AspTyrGluGlySerGlySerAspAlaSerLeuSerSerLeuThrSerSerAlaSer 800
Db 2394 GACTATGAGGCGACGGCTCCGACGCGCGTCCCTGAGCTCCCTCAGCTTCTCGCCTCC 2453
Qy 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGlySerArgPheLysLeuAla 820
Db 2454 GACCAAGACCAAGATTACGATTATCTGAACGAGTGGGCGAGCCGCTTCAAGAACTGGCA 2513
Qy 821 AspMetTyrGlyGlyGlyGluAspAsp 829
Db 2514 GACATGTACGTGGTGGGAGGACGAC 2540
RESULT 3
US-09-880-107-3759
; Sequence 3759, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 3759
; LENGTH: 3171
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 X63629
US-09-880-107-3759
Alignment Scores:
Pred. No.: 0 Length: 3171
Score: 829.00 Matches: 829
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0
US-09-916-849A-1 (1-829) x US-09-880-107-3759 (1-3171)
Qy 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuValCysTrpLeuGln 20
Db 54 ATGGGGCTCCCTCTGAGACCTCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 113
Qy 21 CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
Db 114 TGGCGGCTCCGAGCGGTCCGGGCGGTCTTCAGGAGGCTGAAGTACCTTGAGGCG 173
Qy 41 GlyGlyAlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
Db 174 GGAGCGCGAGCAGGAGCG 233
Qy 61 GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyValThr 80
Db 234 CAAGACCGAGCTCTGTTAGACCTGATATGATGACTTCACTGTGCGGAATGGCGAGCA 293
Qy 81 ValGlnGluArgArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
Db 294 GTCCAGGAAAGAGGTCACTGAGGAAAGGAATCCATTGAAGATCTTCCCATCCAAACGT 353
Qy 101 IleLeuArgArgHisLysLysArgAspTrpValValAlaProIleSerValProGluAsnGly 120
Db 354 ATCTTACGAAGACACAGAGAGATTGGGTGTTCTCTCAATATCTGCTCCCTGAAAAATGGC 413

QY 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
Db 414 AAGGTCCTTCCCTCCAGAGACCTGAATCAGCTCAAGTCTAATAAGATAGAGACACCAAG 473
QY 141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProProGluGlyValPheAlaVal 160
Db 474 ATTTTCTACAGCATCAGCGGGCGGGGAGACAGCCCCCTCAGGGTGCTTCGTGTGA 533
QY 161 GluLysGluThrGlyTyrLeuLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys 180
Db 534 GAGAAGAGAGACAGGCTGTGTGTGTGAATAAGCCACTGACCGGGAGGAGATGGCCAAG 593
QY 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
Db 594 TATGAGCTCTTTGGCCACGCTGTGTGTGAGAAATGGTCCCTCAGTGGAGGCCCATGAAC 653
QY 201 IleSerIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
Db 654 ATCTCCATCATGTCGACCGACCAAGATGACCAAGCCCAAGTTTACCCAGGACACCTTC 713
QY 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
Db 714 CGAGGGAGTGCTTTAGAGGAGTGCTTACCAGGTACTTCTGTGATGAGGTGACAGCCACA 773
QY 241 AspGluAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln 260
Db 774 GATGAGATGATGCCATCATCACTACAAATGGGGTGTGCTTACTCCATCCATAGCCAA 833
QY 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
Db 834 GAACCAAGAGACCCACAGACCTCATGTTCCAAATTCACCGGAGCACAGGCACCATCAGC 893
QY 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
Db 894 GTCATCTCCAGTGGCTGGACCGGGAANAAGTCCCTGAGTACACACTGACCATCCAGGCC 953
QY 301 ThrAspMetAspGlyAspGlySerThrThrAlaValAlaValValGluIleLeuAsp 320
Db 954 ACAGACATGATGGGACGGCTCCACCACCGCAGTGGCAGTGTAGTGGAGATCCTTGAT 1013
QY 321 AlaAsnAspAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
Db 1014 GCCAATGACATGCTCCCATGTTTGACCCCAAGTACGAGGCCCATGTGCTTGAGAT 1073
QY 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
Db 1074 GCAGTGGGCCATGAGGTGCAGAGGCTGACGGTCACTGATCTGCACGCCCCCACTCACC 1133
QY 361 AlaTyrArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
Db 1134 GCGTGGCGTCCACCTACCTTATCATGGCGGTGACGACGGGACCATTTTACCATCACC 1193
QY 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
Db 1194 ACCACCTCTGAGCAACCGAGGCATCTGACACACCGAGAGGTGTGGATTTTGAGGCC 1253
QY 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db 1254 AAAAAACAGACACACCTGTGACGTGAAGTACCAACAGAGCCCTTTTGTGCTGAAGCTC 1313
QY 421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
Db 1314 CCACCTCTCCAGGCCCATAGTGGTCCACGTGAGGATGTGAATGAGGCCCTGTGTTT 1373
QY 441 ValProProSerLysValValGluValGlnGluIleProThrGlyGluProValCys 460
Db 1374 GTCCACCTCCAAAGCTGTGTAGTCCAGAGGGCATCCCACTGGGGAGCCTGTGTGT 1433
QY 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db 1434 GTCTACACTCGAAGACCTCGACAGGGAATCAAAGATACAGCTACCGCATCCTCGAGA 1493
QY 481 AspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500

RESULT 4

US-10-058-270A-125

Db 1494 GACCCAGCAGGGTGCTAGCCATGGACCCAGACACTGGGCAAGTGGAGTGTGGCACC 1553
QY 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db 1554 CTCGACCGTGGAGATGAGCAGTGTGTGAGGAACACATCATATGAAGTATGTGTTGCC 1613
QY 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuThrLeuIleAsp 540
Db 1614 ATGGCAATGGAAAGCCCTCCACCACTGGCAGCGGAACCTTCTGCTAACCATGATTGAT 1673
QY 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 1674 GTCAACGACCATGGGCCAGTCCCTGAGGCCCGTCCAGATCAATCTGCAACCAAGCCCT 1733
QY 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
Db 1734 GTGCGCCACGTCTGAAACATCACGACCAAGACCTGTCTCCCAACACCTCCCTTTCCAG 1793
QY 581 AlaGlnLeuThrAspAspSerAspIleTyrThrAlaGluValAsnGluGluValAsp 600
Db 1794 GCCAGCTCACAGATGACTCAGACATCTACTGACGGCAGGAGGTCAACAGGAGGTGAC 1853
QY 601 ThrValValLeuSerLeuLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
Db 1854 ACAGTGGTCTGTCTCCCTGAAGAAAGTCTCTGAAGCAGGATACATATGACGTGCACCTTCT 1913
QY 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
Db 1914 CTGTCTGACCATGGCAACAAAGAGCAGCTGACGGTGATCAGGGCCACTGTGTGCACATGC 1973
QY 641 HisGlyHisValGluThrCysProGlyProTyrLysGlyGlyPheIleLeuProValLeu 660
Db 1974 CATGGCCATGTGCAAACTGCCCTGGACCTGGAAAGAGAGTTCATCTCCCTGTGCTG 2033
QY 661 GlyAlaValLeuAlaLeuPheLeuLeuValLeuLeuLeuValArgLysLys 680
Db 2034 GGGGCTGTCTGGCTCTGTCTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2093
QY 681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
Db 2094 CGGAGATCAAGGAGCCCTCTCTCTCCAGAGATGACCCCGTGACACGCTCTTCTAC 2153
QY 701 TyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
Db 2154 TATGCGAAGAGGGGGTGGGAGAGCAGGACTATGACATCACCCAGCTCCACCGA 2213
QY 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleLeuPro 740
Db 2214 GGTCTGGAGGCGCAGCGCGAGTGTGTCTCCGCAATGACGTGGCAACCAACCATCATCCG 2273
QY 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
Db 2274 ACACCCATGTACCGTCTTAGGCCAGCCAAACCCAGATGAAATCGGCAACTTTATATTGAG 2333
QY 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
Db 2334 AACCTGAAAGGGCGCTTAACACAGACCCACGCGCCCTACGACACCTCTTGTGTGTC 2393
QY 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
Db 2394 GACTATGAGGCGAGCGGTCCGACCGCGTCCCTGTAGCTCCCTCACCTCTCCGCCCTCC 2453
QY 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTyrGlySerArgPheLysLeuAla 820
Db 2454 GACCAAGACCAAGATTACGATTATCTGAACGAGTGGGCGAGCCGCTTCAAGAAGCTGCA 2513
QY 821 AspMetTyrGlyGlyGluAspAsp 829
Db 2514 GACATGTACGGTGGCGGGGAGGACGAC 2540

654 ATCTCCATCATCGTGACCGACCAAGATGACCAAGCCCAAGTTTATCCAGGACACCTTC 713
Qy
221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
Db
714 CGAGGGAGTGTCTTAGAGGAGTCCCTACCAAGGTACTTCTGTATGAGGTGACAGCCACA 773
Qy
241 AspGluAspAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln 260
Db
774 GATGAGGATGATGCATCTACCTACATCAATGGGGTGGTCTTACTCCATCCATGCCAA 833
Qy
261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
Db
834 GAACCAAGACCCACACAGCCTCATGTTCCAAATTCACCGGACAGGACCATCAGC 893
Qy
281 ValIleSerSerGlyLeuAspArgGluValProGluTyrThrLeuThrIleGlnAla 300
Db
894 GTCAATCTCCAGTGGCTGGACCGGGAAGAAGTCCCTGAGTACACACTGACCATCCAGGCC 953
Qy
301 ThrAspValAspGlyAspGlySerThrThrThrAlaValAlaValValGluIleLeuAsp 320
Db
954 ACAGACATGATGGGACCGCTCCACCACACCGGACGTGGCAGTGGAGATCCTTGAT 1013
Qy
321 AlaAsnAspAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
Db
1014 GCCAATGACAAATGCTCCCAATGTTTACCCCGAGAGTACGAGGCCCATGTGCTGAGAT 1073
Qy
341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
Db
1074 GCATGGGCCATGAGGTGCAGAGGCTCACGGTCACTGATCTGGACGCCCCCAACTCACCA 1133
Qy
361 AlaTrpArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
Db
1134 GCGTGGCGTGCACCTACCTATCATCGGCGGTGACACGGGACCAATTTTACCATCACC 1193
Qy
381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
Db
1194 ACCACCCCTGAGACCAACCGGCTCTGCACCAACGAGGAAGGTTTGGATTGAGGCC 1253
Qy
401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db
1254 AAAAACCCAGCACACCTGTACGTGAGTGAAGTACCAACGAGGCCCTTTTGTGCGAAGCTC 1313
Qy
421 ProThrSerThrAlaThrIleValHisValGluAspValAsnGluAlaProValPhe 440
Db
1314 CCAACCTCCACAGCACCATAGTGTCCAGTGGAGATGTGAATGAGGCACCTGTGT 1373
Qy
441 ValProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
Db
1374 GTCCACCCCTCCAAAGTCGTTAGTCCAGAGGGCATCCCTCCATCGGGAGCGCTGTGT 1433
Qy
461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db
1434 GTCTACACTGCAGAAAGCCCTTCACAGAGAGATCAAAAGATCAGTACCCATCCTTGAGA 1493
Qy
481 AspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db
1494 GAACCCAGCAGGTGGTGGTATGATGACCCAGACAGTGGCGAGGTCAAGCTGTGGGCACC 1553
Qy
501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db
1554 CTCGACCGTGAAGATGAGCAGTGTGTGAGGAACAACATCTATGAAGTCATGGTCTGGGCC 1613
Qy
521 MetAspAsnGlySerProProThrThrThrGlyThrLeuLeuLeuThrLeuIleAsp 540
Db
1614 ATGACCAATGAAGCCCTCCACCATCGGCACGGAAACCTTCTGCTTAACACTGATTGAT 1673
Qy
541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db
1674 GTCAACGACCATGGCCCGTCCCTGAGCCCGTCCAGATCCCATCTGCAACCAAGGCCCT 1733
Qy
561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
Db
1734 GTGGGCCACGTGCTGAACATCAAGGACCTGTCTCCCCACACCTTCCCTTTCCAG 1793

581 AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGlyAsp 600
Db
1794 GCCCAGCTCACAGATGACTCAGACATCTACTGGACGGCAGAGGTCAACGAGGAAGGTGAC 1853
Qy
601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
Db
1854 ACAGTGGTCTTGTCCCTGAAGAAGTTCCTGAAGCAGGATACATATGACGCGACCTTTCT 1913
Qy
621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
Db
1914 CTGCTGACCATGGCAACAGACAGCTGACGCTGATCAGGGCCACTGTGTGCGACTGC 1973
Qy
641 HisGlyHisValGluThrCysProGlyProTyrLysGlyGlyPheIleLeuProValLeu 660
Db
1974 CATGCCCATGTGCAAAACCTGCCCTGGACCCCTGGAAAGAGGTTTCATCCTCCCTGTGCTG 2033
Qy
661 GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysLys 680
Db
2034 GGGGCTGTCTGGCTGCTGCTTCTCTCTGCTGCTGCTGCTTTTGGTGAGAAAGAG 2093
Qy
681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
Db
2094 CGGAAGATCAAGGAGCCCTCTACTCTCCAGAGATGACACCCCTGACAACTGCTTCTAC 2153
Qy
701 TyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
Db
2154 TATGCGAAGAGGGGGTGGCGAAGAGACGAGACTATGACATCACCCAGCTCCACCGA 2213
Qy
721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
Db
2214 GGTCTGAGGCGCAGCGCGAGGTGTTCTCCGCAATGACGTGGCACCACCATCATCCCG 2273
Qy
741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
Db
2274 ACACCCATGTACCTGCTTAGCCAGCAGCAACCATGAATATCGGCACTTTATTAATTGAG 2333
Qy
761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
Db
2334 AACCTGAGGCGGCTACACAGACCCACAGCCCGCCCTACGACACCTCTTGGTGTTC 2393
Qy
781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
Db
2394 GACTATGAGGCGCGGCTCCGACCGCGCTGCTGAGCTCCCTCAGCTCCTCCGCTCC 2453
Qy
801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGlySerArgPheLysLeuAla 820
Db
2454 GACCAAGACCAAGATTACGATTATCTGACGATGGGGCAGCGCTTCAAGAACTGGCA 2513
Qy
821 AspMetTyrGlyGlyGluAspAsp 829
Db
2514 GACATGTACGCTGGCGGGGAGGACGAC 2540

RESULT 6

US-10-172-118-669
; Sequence 669, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
; CURRENT APPLICATION NUMBER: US/10/172,118
; PRIOR FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/380, 770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699

Db 1194 ACCACCTGAGAGCAACCGAGCATCTCTGACAAACAGGAGGTTTGGATTTTGGAGCC 1253
Qy 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db 1254 AAAAACAGCAGCACACCTGTACGTTGAAGTACCAACAGAGGCCCTTTTGTGCTGAAGCTC 1313
Qy 421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
Db 1314 CCAACCTCCACAGCCACCATAGTGGTCCACGTCGAGGATGTGAATGAGGACCTGTGT 1373
Qy 441 ValProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
Db 1374 GTCCCACTCCAAAGTCGTTGAGTCCAGGAGGCATCCCCACTGGGAGGCTGTGT 1433
Qy 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db 1434 GTCTACACTGCAGAGACCCCTGCAAGGAGAAATCAAAGATCAGCTACCGCATCTCTGAGA 1493
Qy 481 AspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db 1494 GACCCAGCAGGCTGGCTAGCCATGGACCAAGCAGTGGCAGGTCCACACTGTGGCACC 1553
Qy 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db 1554 CTGCGACGTCAGGATGAGCAGTGTGTGAGGAAACAACATCTATCAAGTCATGCTTTGGCC 1613
Qy 521 MetAspAsnGlySerProThrThrGlyThrGlyThrLeuLeuLeuThrIleLeuAsp 540
Db 1614 ATGGAATATGGAAGCCCTCCCACTGGCAGCGGAACCTCTCTGCTAATCACTGATTGAT 1673
Qy 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 1674 GTCAACGACCATGGCCAGTCCTCTGAGCCCGTCAGATCACCATCTGCAACCAAGCCCT 1733
Qy 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
Db 1734 GTGCGCCAGCTGCTGAACATCAGGCAAGACCTGTCTCCCAACACCTCCCTCTTCAG 1793
Qy 581 AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGlyAsp 600
Db 1794 GCGCAGCTCACAGATGACTCAGACATCTACTGGACGGCAGAGGTCAACGAGAAAGTGAC 1853
Qy 601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
Db 1854 ACAGTGTCTTGTCTCCCTGAAGAAGTCTCTGAAGCAGGATACATATCAGTCGACCTTCT 1913
Qy 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
Db 1914 CTGCTGACCATGGCAACAAAGACGACGTGACGGTGTATCAGGGCCACTGTGTGCGACTGC 1973
Qy 641 HisGlyHisValGluThrCysProGlyProTyrLysGlyPheIleLeuProValLeu 660
Db 1974 CATGGCCATGTGAAAACCTGCCTGGACCTGGAAAGGAGGTTTCATCTCCCTGTGCTG 2033
Qy 661 GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuValLeuLeuValArgLysLys 680
Db 2034 GGGCGTCTCTGTCTGTCTGTCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2093
Qy 681 ArgLysIleLysGluProLeuLeuProGluAspThrArgAspAsnValPheTyr 700
Db 2094 CGGAAGATCAAGAGCCCTCTACTCCAGAAAGATGACACCCGTCGACAACTCTTCTAC 2153
Qy 701 TyrGlyGluGlyGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
Db 2154 TATGGCGAAGAGGGGTGGCGAAGAGGACAGGACTATGACATCACCAGCTCCACCGA 2213
Qy 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleLeuPro 740
Db 2214 GGTCTGGAGCCAGGCCGAGGTGTCTTCCCAATGACGTGGCAACCAACCATATCCCG 2273
Qy 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleLeuGlu 760

Db 2274 ACACCCATGTACGTCCTAGCCAGCCAAACCAGATGAAATCGGCACACTTTATTAATTGAG 2333
Qy 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProTyrAspThrLeuLeuValPhe 780
Db 2334 AACCTGAAGCGGTATACACAGACCCACAGCCCGCCCTACGACACCTCTTGGTGTTC 2393
Qy 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
Db 2394 GACTATGAGGACGCGGTCCGACGCGCGTCCCTGAGTCCCTCAGCTCCTCCGCTCC 2453
Qy 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGlySerArgPheLysLeuAla 820
Db 2454 GACCAAGACCAAGATTACGATTATCTGACGAGTGGGCGACCCGCTTCAAGAAGCTGGCA 2513
Qy 821 AspMetTyrGlyGlyGlyGluAspAsp 829
Db 2514 CACATGTACGTTGGGGGAGGACGAC 2540
RESULT 8
US-10-158-123-2
; Sequence 2, Application US/10158123
; Publication No. US2003019406A1
; GENERAL INFORMATION:
; APPLICANT: Reinhard, Christoph
; APPLICANT: Klingner, Julie
; APPLICANT: Jefferson, Ann
; APPLICANT: Escobedo, Jaime
; APPLICANT: Randazzo, Filippo
; APPLICANT: Winter, Jill
; APPLICANT: Goodson, Robert
; APPLICANT: Qi, Weimin
; TITLE OF INVENTION: P-Cadherin as a Target for Anti-Cancer
; FILE REFERENCE: 35784/259994
; CURRENT APPLICATION NUMBER: US/10/158,123
; PRIOR FILING DATE: 2003-02-07
; PRIOR APPLICATION NUMBER: 60/294,225
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 3171
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-158-123-2
Alignment Scores:
Pred. No.: 0 Length: 3171
Score: 829.00 Matches: 829
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 15 Gaps: 0
US-09-916-849A-1 (1-829) x US-10-158-123-2 (1-3171)
Qy 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuGlnValCysTrpLeuGln 20
Db 54 ATGGGGCTCCCTCGTCGACCTCTCGCTCTCTCTCTCCAGGTTTGTGCTGCTGAG 113
Qy 21 CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
Db 114 TGCAGCCCTCCGAGCGCTGCGGCGGTCTTTCAGGAGGCTGAAGTGACCTTGGAGGCG 173
Qy 41 GlyGlyAlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
Db 174 GGAGCGCGGAGCAGGAGCGCCGACGCTGGGAAAGTATTCATGGGCTGCCCTGGG 233
Qy 61 GlnGluProAlaLeuPheSerThrAspAsnAspAspPheThrValArgAsnGlyGluThr 80
Db 234 CAAGAGCAGCTCTGTGTAGCACTGATAATGATGACTTCATGTGCGGAATGGCAGACA 293
Qy 81 ValGlnGluArgArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100

Db 294 GTCCAGGAAGAGGTCACCTGAAGGAAGAAATCCATTCAGATCTTCCATCCAAACGT 353
Qy 101 ILeuAArgHISlyArgAspTTPValAlaProIleSerValProGluAsnGly 120
Db 354 ATCTTACGAGACACAGAGAGATGGGTGGTTCCTCAATATCTGTCTGCTGAAATGGC 413
Qy 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
Db 414 AAGGGTCCCTTCCCCAGAGACTGATCAGCTCAAGTCTAATAAGATAGACACCAAG 473
Qy 141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProProGluGlyValPheAlaVal 160
Db 474 ATTTCTACAGCATCAGGGGCGGGGCGAGACAGCCCTGAGGGTGTCTTCGCTGTA 533
Qy 161 GluLysGluThrGlyTTPLeuLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys 180
Db 534 GAGAAGGAGACAGCTGGTGTGGTGGATTAAGCCACTGGACCGGGAGAGATTGCCAAG 593
Qy 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
Db 594 TATGAGCTCTTTGGCCACGCTGTGTGAGAGATGGTGCCTCAGTGGAGGACCCCATGAAC 653
Qy 201 IleSerIleIleValThrAspGlnAsnAspHisLysProIysPheThrGlnAspThrPhe 220
Db 654 ATCTCCATCATCTGGACCGACCAAGATGACCAAGCCCAAGTTTACCAGGACACCTTC 713
Qy 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
Db 714 CGAGGGAGTGTCTAGAGGGAGTCTTACCAGGTACTTCTGTATGTCAGGTGACAGCCACA 773
Qy 241 AspGluAspAlaIleTyrThrAsnGlyValAlaValAlaTyrSerIleHisSerGln 260
Db 774 GATGAGGATGATGCCATCTACACTCAATGGGGTGGTGTCTTACTCCATTCATAGCCAA 833
Qy 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
Db 834 GATCCAAAGAGACCCACACGACCTTCACTGTTTCACTTCCAGGACAGCACCACATCAGC 893
Qy 281 ValIleSerSerGlyLeuAspAArgLysValProGluTyrThrLeuThrIleGlnAla 300
Db 894 GTCATCTCCAGTGGCTGGACCGGAAAAAGTCCCTGAGTACACACTGACCATCCAGGCC 953
Qy 301 ThrAspMetAspGlyAspGlySerThrThrAlaValAlaValValGluIleLeuAsp 320
Db 954 ACAGACATGATGGGACGGCTCCACACCGGAGTGGCAGTAGTGGAGATCTTGTAT 1013
Qy 321 AlaAsnAspAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
Db 1014 GCCAATGACAATGTCTCCATGTTTGAACCCCAAGATACGAGGCCCATGTGCTGAGAAAT 1073
Qy 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
Db 1074 GCAGTGGGCGCATGAGGTGCAGAGCTGACGGTCACTGATCTGACGCCCCCAACTCACCA 1133
Qy 361 AlaTTPArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
Db 1134 GCGTGGGTGTCACCTACTATCATGGGCGGTGACGACCGGGACCATTTTACCATCACC 1193
Qy 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
Db 1194 ACCCACCTGAGAGCAACCGAGGCATCTGACAAACAGGAGGGTTTGGATTTTGAGGCC 1253
Qy 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db 1254 AAAAACCCAGACACCCCTGACCTTGAAGTGAACCAACGAGGCCCTTTTGTGCTGAAGCTC 1313
Qy 421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
Db 1314 CCAACCTCCACAGCCACCATAGTGGTCCAGTGGAGGATGTGAATGAGGCACCTGTGT 1373
Qy 441 ValProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460

Db 1374 GTCCCACTCCAAAGTCGTGTGAGGTCCAGAGGGCATCCCCACTGGGAGACCTGTGTGT 1433
Qy 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db 1434 GTCTACACTGACAGAGACCTCCAGAGGAGATCAAAAGATCAGTACCGCATCTGAGA 1493
Qy 481 AspProIaGlyTTPLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db 1494 GACCAGCAGGGTGGCTAGCATGGACCCAGACAGTGGGAGGTCCAGCTGTGGGCACC 1553
Qy 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db 1554 CTCACCGTGGAGTATGACAGTGTGTGAGGACACACATCTATGAAGTCAATGGCTTGGCC 1613
Qy 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuLeuAsp 540
Db 1614 ATGGACAATGGAAGCCCTCCCACTGGCACGCGGAACCTTCTGTAACTGATTTGAT 1673
Qy 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 1674 GTCAACGACCATGGCCAGTCCCTGAGCCCCGTCAGATCACCATCTGCAACCAAGCCT 1733
Qy 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
Db 1734 GTGGCCACAGTGTGAACATCAGGACAGACCTGTCTCCCACTCCCTCCCTTTCCAG 1793
Qy 591 AlaglnLeuThrAspAspSerAspIleTyrTTPThrAlaGluValAsnGluGluValAsp 600
Db 1794 GCCAGCTCAGATGACTCAGACATCTCTGGACGCGAGGTCAACGAGGAAGGTGAC 1853
Qy 601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
Db 1854 ACAGTGTCTTGTCCCTGAGAAGTTCCTGAGCAGCATACATATGACGTGACCTTTCT 1913
Qy 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
Db 1914 CTGTCTCACCATGGCAAAAGAGCAGCTACGGTGTATCAGGGCCACTGTGTGGACTGC 1973
Qy 641 HisGlyHisValGluThrCysProGlyProTTPLysGlyGlyPheIleLeuProValLeu 660
Db 1974 CATGCCATGTCCAAACTGCCCTGGACCTGGAAAGAGGTTCATCCCTCCCTGTGCTG 2033
Qy 661 GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysLys 680
Db 2034 GGGGCTGTCTGGTCTGTCTGTCTCTGTGTGTGTCTTTTGTGTGTGAGAAAGAG 2093
Qy 681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
Db 2094 CGGAAGATCAAGGAGCCCTCTCTACTCCAGAGATGACCCCGTGACACGCTTCTAC 2153
Qy 701 TyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
Db 2154 TATGGGAGAGGGGGGTGGCGAAGAGACCCAGCACTATGACATCACCCAGTCCACCGA 2213
Qy 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
Db 2214 GGTCTGGAGGCCAGCGCGAGGTGTCTCCGAATGACGTGGGACCAACCATCATCCCG 2273
Qy 741 ThrProMetTyrAspProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
Db 2274 ACACCCATGTACCTCTAGGCCAGCAACCCAGATGAATCGGCAACTTATTAATTTAG 2333
Qy 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
Db 2334 AACCTGAAGGGGGTAAACACAGACCCACGCCGCCCTTACGACACCTCTTGGTGTTC 2393
Qy 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
Db 2394 GACTATGAGGCGGGCTCCGACGCCGCTCCCTGAGCTCCCTCACCTCTCCGCTCC 2453
Qy 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTTPGlySerArgPheLysLeuAla 820
Db 2454 GACCAAGACCAAGATTACGATTATCTGAACGAGTGGGCGAGCCGCTTCAAGAGACTGGCA 2513

Db 1254 AAAAACACGACACCCCTGTACGTTGAAGTGACCAACGAGGCCCTTTTGTGCTGAAGCTC 1313
Qy 421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
Db 1314 CCAACCTCCACAGCCACCATAGTGGTCCACGTGGAGGATGTGAATGAGGACCTGTGTTT 1373
Qy 441 ValProProSerLysValValGluValGlnGluGlyLeuProThrGlyGluProValCys 460
Db 1374 GTCCACCCCTCCAAAGTCGTTGAGTCCAGAGGCGATCCCACTGGGAGCGCTGTGT 1433
Qy 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db 1434 GTCTACACTGCAGAGACCCCTGCACAGAGAAATCAAAAGATCAGCTACCGCATCTTGAGA 1493
Qy 481 AspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db 1494 GACCCAGCAGGGTGGCTAGCATGGACCCAGACAGTGGCGAGGTACAGCTGTGGGCACC 1553
Qy 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db 1554 CTCGACCGTGAGGATGAGCAGTTTGTGAGGAACAACATCTATGAAGTCATGTGTTGGCC 1613
Qy 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuLeuAsp 540
Db 1614 ATGACAAATGGAAGCCCTCCACACACTGGCAGGGAACCCCTTCTGCTAACACTGATGAT 1673
Qy 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrLeuCysAsnGlnSerPro 560
Db 1674 GTCAACGACCATGCCCCAGTCCCTGAGCCCGTCAGATCACTCATCTGCAACCAAGGCCCT 1733
Qy 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
Db 1734 GTGGCCACCGTGGTGAACATCACGGAAGAGCCCTGTCTCCACACCTCCCTTTTCCAG 1793
Qy 581 AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGlyAsp 600
Db 1794 GCCAGCTCACAGATGACTCAGACATCTACTGGAGCGCAGAGGTCAACGAGGAAGTGAC 1853
Qy 601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
Db 1854 ACAGTGTCTGTGCTCCCTGAAGAGTCTCTGAAGCAGGATACATATGACGTGCACCTTCT 1913
Qy 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
Db 1914 CTGTCTGACCATGCGAACAACAGACAGCTGACGTGTATCAGGSCCATCTGTGGACTGC 1973
Qy 641 HisGlyHisValGluThrCysProGlyProThrLysGlyPheIleLeuProValLeu 660
Db 1974 CATGCCATGTCCAAACCTGCGCTGGACCTCGAAGAGGAGGTTTTCATCTCCCTGTGCTG 2033
Qy 661 GlyAlaValLeuAlaLeuLysPheLeuLeuValLeuLeuLeuValAlaGlyLysLys 680
Db 2034 GGGGCTGTCTGGCT 2093
Qy 681 ArgLysIleLysGluProLeuLeuLeuProGluAspThrArgAspAsnValPheTyr 700
Db 2094 CGGAAGATCAAGAGGCCCTCTACTCCCAAGATGACACCCGTGACAACGCTCTTCTAC 2153
Qy 701 TyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
Db 2154 TATGGCGAAGAGGGGGTGGGAGAGAGACCCAGGACTATGATCATCCAGCTCCACCGA 2213
Qy 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIlePro 740
Db 2214 GGTCTGGAGGCCAGCCCGAGGTGTCTTCGCAATGACGTGGCACCACCAACCATCATCCG 2273
Qy 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleGlu 760
Db 2274 ACACCCATGTACCGTCTAGGCGAGCCCAACCCAGATGAAATCGGCAACTTTATTAATGAG 2333
Qy 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProThrAlaProThrLeuLeuValPhe 780
Db 2334 AACCTGAAGGGGGCTTAACACAGACCCCAACAGCCCGCCCTACGACACCTCTTGTGTTCT 2393

Qy 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
Db 2394 GACTATGAGGCGAGCGGCTCCGACCGCGCTCCCTGAGCTCCCTCAGCTCCCTCCGCTCC 2453
Qy 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGlySerArgPheLysLysLeuAla 820
Db 2454 GACCAAGACCAAGATTAGATTATCTGAACGATGGGCGAGCCGCTTCACAGAGCTGGCA 2513
Qy 821 AspMetTyrGlyGlyGlyGluAspAsp 829
Db 2514 GACATGTACGTTGGCGGAGGACGAC 2540
RESULT 10
US-10-295-027-799
; Sequence 799, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; FILE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 799
; LENGTH: 3171
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-295-027-799
Alignment Scores:
Pred. No.: 0 Length: 3171
Score: 829.00 Matches: 829
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 16 Gaps: 0
US-09-916-849A-1 (1-829) x US-10-295-027-799 (1-3171)
Qy 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuGlnValCysTrpLeuGln 20

Db	54	ATGGGGCTCCCTCGTGGACCTCTCGCGTCTCTCCCTCTCCAGGTTGCTGGCTGCAG	113
Qy	21	CysAlaalaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla	40
Db	114	TGCGCGGCTCCGAGCGTCCGGCGGTCCTCAGGAGGCTGAAGTGCACCTTGGAGCG	173
Qy	41	GlyGlyAlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly	60
Db	174	GGAGCGGGAGCAGGAGCCCGCGCAGCGCTGGGGAAAGTATTTCATGGGCTGCCCTGG	233
Qy	61	GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr	80
Db	234	CAAGAGCCAGCTCTGTTAGCACTGATAATGATGACTTCATGTGCGGAAATGGCGAGAC	293
Qy	81	ValGlnGluArgArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg	100
Db	294	GTCAGAGAAAGAGGTCACTGAAGGAAAGGAATCCATTGAAGATCTTCCCATCCAAACGT	353
Qy	101	IleLeuArgArgHisLysArgTyrValValAlaProIleSerValProGluAsnGly	120
Db	354	ATCTTAGAGACACACAGAGAGATTGGGTGGTTGCTCCAAATATCTGTCCTGAAATGGC	413
Qy	121	LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys	140
Db	414	AAGGGTCCCTTCCCCAGAGACTGAATCAGCTCAAGTCTAATATAAGATAGAGACACCAAG	473
Qy	141	IlePheTyrSerIleThrGlyProGlyAlaAspSerProProGluGlyValPheAlaVal	160
Db	474	ATTTCACAGCATCACGGGCGCGGGGCACAGACCCCTCAGTGGAGGTCTTCGCTGTGA	533
Qy	161	GluLysGluThrGlyTyrProLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys	180
Db	534	GAGAAAGAGACAGCGCTGTTGTTGTAATAAGCCACTGGACCGGAGGAGATTGCCAAG	593
Qy	181	TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn	200
Db	594	TATGAGCTCTTTGGCCACGCTGTGTACAGAAATGGTGCTCAGTGGAGGACCCCATGAAC	653
Qy	201	IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe	220
Db	654	ATCTCCATCATGTGACCGACCAAGATGACCACAGGCCAAGTTTACCAGGACACCTTC	713
Qy	221	ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr	240
Db	714	CGAGGAGTGTCTTAGAGGAGTCCCTACCAGGTACTTCTGTGATGAGTGCAGGACCCACA	773
Qy	241	AspGluAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln	260
Db	774	GATGAGGATGATGCCATCTACACCTCAATGGGTGGTTGCTTACTCCATCCATAGCCAA	833
Qy	261	GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer	280
Db	834	GAACCAAGGACCCACACAGCCTCATGTTCAATTCACCGGACACAGGCACCATCAGC	893
Qy	281	ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGluAla	300
Db	894	GTCAATCCAGTGGCTGGACCGGGAAAAAGTCCCTGAGTACACACTGCACCATCAGGCC	953
Qy	301	ThrAspMetAspGlyAspGlySerThrThrThrAlaValAlaValGluIleLeuAsp	320
Db	954	ACAGACATGATGGGACGGCTCCACCACCACCGCAGTGGCAGTAGTGGAGATCCTTGAT	1013
Qy	321	AlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn	340
Db	1014	GCCAAATGACAAATGCTCCCATCTTTTGACCCCCAGAAAGTACGAGGGCCCATGTGCCT	1073
Qy	341	AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro	360
Db	1074	CGAGTGGGCCATGASGTGCAGAGGCTGACGGTCACTGATCTGGACGCCCCCACTACCA	1133
Qy	361	AlaTyrArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr	380

Db	1134	CGTGGCGTGCACCTACCTATTATCATGGCGGTGCACACGGGACCAATTTTACATCACCC	1193
Qy	381	ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla	400
Db	1194	ACCACCTGAGAGCAACACGGGCATCCTGACACACAGAGAGGTGTGGATTGTAGGCC	1253
Qy	401	LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu	420
Db	1254	AAAAACACAGACACCCCTGTAGTTGAAGTGACCAACAGAGCCCTTTGTGCTGAAGCTC	1313
Qy	421	ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe	440
Db	1314	CCAACTCTCCACAGCACCATAGTGGTCCACGTGGAGGATGTGAATGAGGCACCTGTGTT	1373
Qy	441	ValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys	460
Db	1374	GTCCACCCCTCCAAAGTCGTTGAGTCCAGAGGACATCCCCATCGGGAGCCTGTGTGT	1433
Qy	461	ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg	480
Db	1434	GTCTACTCGAGAGAACCTGTACAGAGAGATCAAAAGATCAGTACCCGATCCTTGAGA	1493
Qy	481	AspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr	500
Db	1494	GACCCAGCGGTGGCTAGCCATGGACCCACAGACAGTGGCGAGGTCA CAGCTGTGGGCACC	1553
Qy	501	LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla	520
Db	1554	CTCAGCCGTGAGGATGAGCTAGTTGTGAGGAAACAATCTATGAAGTCATGGTCTTGGCC	1613
Qy	521	MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAsp	540
Db	1614	ATGGACAATGGAAGCCCTCCACCACTGGCACGGGAACCCCTTCTGCTAACACTGATTGAT	1673
Qy	541	ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro	560
Db	1674	GTCAACACCATGCCCCAGTCCCTGAGCCCGCTCAGATCACCATCTGTCAACCAAGGCCCT	1733
Qy	561	ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln	580
Db	1734	GTGGCCACACGTGTGAACATCACGGACAAAGACCTGTCTCTCCACACCTCCCTTCCAG	1793
Qy	581	AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGlyAsp	600
Db	1794	GCCCAGCTCACAGATGACTCAGACATCTACTGACGCGCAGAGGTCAACGAGGAAGGTGAC	1853
Qy	601	ThrValValLeuSerLeuLysPheLeuLysGlnAspThrTyrAspValHisLeuSer	620
Db	1854	ACAGTGTCTTGTCCCTGAAGAAGTTCCTGAAGCAGGATACATATGACGTGCACCTTCT	1913
Qy	621	LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys	640
Db	1914	CTGTCTCACCATGGCAACAAAGAGCAGCTGACGGTGTATCAGGGCCACTGTGTCCGACTGC	1973
Qy	641	HisGlyHisValGluThrCysProGlyProTrpLysGlyGlyPheIleLeuProValLeu	660
Db	1974	CATGGCCATGTCAAAACCTCCCTGACCCCTGAAAGGAGTTTCATCTCCCTGTGCTG	2033
Qy	661	GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysLys	680
Db	2034	GGGCTCTCTGGCTCTGCTGTTCCCTCTGCTGCTGCTTTTGTGTGAGAGAAGAG	2093
Qy	681	ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr	700
Db	2094	CGGAAGATCAAGAGGCCCTCTTACTCCCAAGAAGATGACACCGGTGACACGCTTCTTAC	2153
Qy	701	TyrGlyGluGluGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisArg	720
Db	2154	TATGGCGAAGAGGGGGTGGCGAAGAGGACCAGCATATGACATACACCCAGCTCCACCGA	2213
Qy	721	GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleLeuPro	740
Db	2214	GGTCTGAGCGCCAGCCGAGGTGGTTCTCCGAATGACGTGGCAACAACCAATCATCCCG	2273

QY 741 ThrProMetTyrArgProArgProAlaAsnProAspGluLeuGlyAsnPheLeuGlu 760
 Db 2274 ACACCAATGACCGTCTAGCCAGCCACCCAGATGAATCGCAACTTATATTGAG 2333
 QY 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
 Db 2334 AACCTGAAGCGGCTAAACACAGACCCACAGCCCGCCCTACGACACCCCTCTTTGGTGTTC 2393
 QY 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
 Db 2394 GACTATGAGGCGACGGCTCCGAGCCCGCTCCCTGAGCTTCTTACCTCTCCGCTCC 2453
 QY 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGlySerArgPheLysLeuAla 820
 Db 2454 GACCAAGACCAAGATTACGATTATCTACAGAGTGGGCGAGCGCTTCAAGAGCTGGCA 2513
 QY 821 AspMetTyrGlyGlyGlyGluAspAsp 829
 Db 2514 GACATGTACGGTGGCGGGAGGAGCAG 2540

RESULT 11

US-10-295-027-862
 ; Sequence 862, Application US/10295027
 ; Publication No. US20030232350A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Afar, Daniel
 ; APPLICANT: Aziz, Natasha
 ; APPLICANT: Ginsberg, Wendy M.
 ; APPLICANT: Gish, Kurt C.
 ; APPLICANT: Glyne, Richard
 ; APPLICANT: Hevezi, Peter A.
 ; APPLICANT: Mack, David H.
 ; APPLICANT: Murray, Richard
 ; APPLICANT: Watson, Susan R.
 ; APPLICANT: Eos Biotechnology, Inc.
 ; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
 ; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
 ; FILE REFERENCE: 018501-012500US
 ; CURRENT APPLICATION NUMBER: US/10/295,027
 ; CURRENT FILING DATE: 2002-11-13
 ; PRIOR APPLICATION NUMBER: US 09/663,733
 ; PRIOR FILING DATE: 2000-09-15
 ; PRIOR APPLICATION NUMBER: US 60/350,666
 ; PRIOR FILING DATE: 2001-11-13
 ; PRIOR APPLICATION NUMBER: US 60/335,394
 ; PRIOR FILING DATE: 2001-11-15
 ; PRIOR APPLICATION NUMBER: US 60/332,464
 ; PRIOR FILING DATE: 2001-11-21
 ; PRIOR APPLICATION NUMBER: US 60/334,393
 ; PRIOR FILING DATE: 2001-11-29
 ; PRIOR APPLICATION NUMBER: US 60/340,376
 ; PRIOR FILING DATE: 2001-12-14
 ; PRIOR APPLICATION NUMBER: US 60/347,211
 ; PRIOR FILING DATE: 2002-01-08
 ; PRIOR APPLICATION NUMBER: US 60/347,349
 ; PRIOR FILING DATE: 2002-01-10
 ; PRIOR APPLICATION NUMBER: US 60/355,250
 ; PRIOR FILING DATE: 2002-02-08
 ; PRIOR APPLICATION NUMBER: US 60/356,714
 ; PRIOR FILING DATE: 2002-02-13
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 1386
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 862
 ; LENGTH: 3171
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-10-295-027-862

Alignment Scores:
 Pred. No.: 0 Length: 3171
 Score: 829 Matches:

Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 16 Gaps: 0
 US-09-916-849A-1 (1-829) x US-10-295-027-862 (1-3171)
 QY 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuGlnValCysTrpLeuGln 20
 Db 54 ATGGGGCTCCCTCGTGGACCTCTCGCGTCTCTCTCTCTCTCCAGGTTTGTGCTCGAG 113
 QY 21 CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
 Db 114 TCGCGGCTCCGAGCGCTGCGGCGGTCTTCAGGAGGCTGAGTGACCTTCGAGGCG 173
 QY 41 GlyGlyAlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
 Db 174 GGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 233
 QY 61 GlnGluProAlaLeuPheSerThrAspAspAspPheThrValArgAsnGlyGluThr 80
 Db 234 CAAGAGCCAGCTCTCTTTAGCACTGATTAATGATGACTTCACTGTGCGGATGGCGACA 293
 QY 81 ValGlnGluArgArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
 Db 294 GTCCAGGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 353
 QY 101 IleLeuArgArgHisLysArgAspTrpValValAlaProIleSerValProGluAsnGly 120
 Db 354 ATCTTTACGAGACACAGAGAGATGGGTGTTCTCTCAATATCTGCTCCCTGAAATGGC 413
 QY 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
 Db 414 AAGGTCTCTTCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 473
 QY 141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProProGluGlyValPheAlaVal 160
 Db 474 ATTTTCTACAGCATCACGGGGCGGGGCGAGACAGCCCCCTGAGGGTGTCTTCGCTGTA 533
 QY 161 GluLysGluThrGlyTrpLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys 180
 Db 534 GAGAAGGAGACAGGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 593
 QY 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
 Db 594 TATGAGCTTTTGGCCACGCTGTGTGAGAGAATGGTGTCTCAGTGGAGGAGCCCATGAAC 653
 QY 201 IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
 Db 654 ATCTCCATCATGTCGTGACCGACAGATGACCAAGCCCAAGTTTACCAGGACACCTTC 713
 QY 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
 Db 714 CGAGGAGGTGCTTAGAGGGAGTCTTACAGGTACTTCTGTGATGAGGTGACAGCCACA 773
 QY 241 AspGluAspAspAlaIleTyrThrAsnGlyValValAlaTyrSerIleHisSerGln 260
 Db 774 GATGAGGATGATGCCATCATCCTTACCAATGGGTGTTGCTTACTCATCCATAGCCAA 833
 QY 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
 Db 834 GACCAAGAGCCACACAGCTCATGTTCACATTCACCGGAGCACAGGACCATCAGC 893
 QY 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGluAla 300
 Db 894 GTCACTCCAGTGGCTCGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 953
 QY 301 ThrAspMetAspGlyAspGlySerThrThrAlaValAlaValGluIleLeuAsp 320
 Db 954 ACAGACATGGATGGGACGGCTCCACACCGGAGTGGCAGTAGTAGTGGAGATCTCTGAT 1013
 QY 321 AlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340

Db 1014 GCCAATGACAATGCTCCCATGTTTGAACCCAGAGTACGAGGCCCATGTGCTCGAAT 1073
Qy 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
Db 1074 GCATGGCCCATAGGTGACAGGCTACGGCTACTGATCTGAGCGCCCACTACCA 1133
Qy 361 AlaTyrArgAlaThrTyrLeuIleMetGlyValAspAspGlyAspHisPheThrIleThr 380
Db 1134 GCGTGGCGTGCACCTACCTTATCATGGCGGTGACGACGGGACCACTTTTACCATCAC 1193
Qy 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgIysGlyLeuAspPheGluAla 400
Db 1194 ACCACCCCTGAGACCAACCGAGGCTCTGACACACCGAGAGGGTTGGATTGTGAGGCC 1253
Qy 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db 1254 AAAAAACGACACACCTGTAGTGAAGTACCAACGAGGCCCTTTTGTGCTGAAGCTC 1313
Qy 421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
Db 1314 CCAACCTCCACAGCCACCATAGTGTCCAGCTGAGGATGTGAATGAGGACCTGTGTTT 1373
Qy 441 ValProProSerLysValValGluValGlnGlyIleProThrGlyGluProValCys 460
Db 1374 GTCCACCTCCAAAGTCGTGTAGTCCAGAGGCGCATCCCACTGGGAGCCCTGTGTGT 1433
Qy 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db 1434 GTCTACACTCAGAAAGACCTGACAGAGAGATCAAAAGATCAGTACCGCATCTTGAGA 1493
Qy 481 AspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db 1494 GACCCACAGGGTGGCTAGCATCGACCCAGAGAGTGGGAGGTGACAGTGTGGGCACC 1553
Qy 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db 1554 CTCGACCGTGAGGATGAGCAGTTTGTGAGGACAAACATCATGATGATGCTTGGCC 1613
Qy 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAsp 540
Db 1614 ATGGACATCGAAGCCCTCCACCATCGACCTGGCAGCGGAAACCTTCTGCTAAACACTGAT 1673
Qy 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 1674 GTCAACGACCATGGCCAGCTCCCTGAGCCCGTCAGATCACTATCGACCAACCAAGCCCT 1733
Qy 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
Db 1734 GTGCGCCACGTGTGAACATCAGGACAGGACCTGTCTCCCAACACCTCCCTTCCAG 1793
Qy 581 AlaGlnLeuThrAspAspSerAspIleTyrThrAlaGluValAsnGluGlyAsp 600
Db 1794 GCCAGCTCAGATGACTCAGACATCTCTGACGCGCAGAGTCAACGAGGAAGTGAC 1853
Qy 601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
Db 1854 ACAGTGTCTTGTCCCTGGAAGAGTTCTCTGACGAGGATACATATGACGTGCACCTTTCT 1913
Qy 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
Db 1914 CTGTCTACCATGGCAACAAAGACGAGTCAACGCTGATCAGGGCCACTGTGTGCGACTGC 1973
Qy 641 HisGlyHisValGluThrCysProGlyProTyrLysGlyGlyPheIleLeuProValLeu 660
Db 1974 CATGGCCATGCAACCTGCCCTGACCTCGAAGAGGTTTCATCTCCCTCTGTGCTG 2033
Qy 661 GlyAlaValLeuAlaLeuPheLeuLeuValLeuLeuValLeuValArgLysLys 680
Db 2034 GGGGCTGTCTGCTGTCT 2093
Qy 681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
Db 2094 CGGAAGATCAAGAGGCCCT 2153

Qy 701 TyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
Db 2154 TATGGCGAAGAGGGGGTGGCGAAGAGGACCACTATGACATCACCAGCTCCACCGA 2213
Qy 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
Db 2214 GGTCTGGAGGCCAGCGCGAGGTGTCTCCGCAATGAGCTGGCACCACCATCATCCG 2273
Qy 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
Db 2274 ACACCCATGTACCGCTCTAGGCCAGCCAAACCCAGATGAATCGCAACTTTATAATTGAG 2333
Qy 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProTyrAspThrLeuLeuValPhe 780
Db 2334 AACCTGAAGCGGCTAAACACAGACCCACAGCCCGCCCTACGACACCTCTTGTGTTC 2393
Qy 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
Db 2394 GACTATGAGGCGAGCGGTCCGACCGCGTCCCTGAGCTCCCTCACCTCTCCCGCTCC 2453
Qy 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTyrGlySerArgPheLysLysLeuAla 820
Db 2454 GACCAAGACCAAGATTACATTATCTGAACGAGTGGGAGCCGCTTCAAGAAGCTGGCA 2513
Qy 821 AspMetTyrGlyGlyGlyGluAspAsp 829
Db 2514 GACATGTACGTTGGCGGGGAGGAGCAG 2540

RESULT 12

US-10-295-027-895
; Sequence 895, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of diagnosis of Cancer, Compositions and
; FILE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 895

354 ATCTTACGAGACACAGAGAGATTGGGTGGTTGCTCCCAATATCTGTCTCCCTGAAATGGC 413
121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
414 AAGGGTCCCTTCCCCAGAGACTGAATCAGCTCAAGTCTAATAAGATAGAGACACCAAG 473
141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProGluGlyValPheAlaVal 160
474 ATTTTCTACAGCAATCAGGGGCCGGGGCGACAGCCCCCTGAGGGTGTCTTCGTGTA 533
161 GluLysGluThrGlyTyrLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys 180
534 GAGAAGGAGACAGAGCTGTGTGTGTAATAAGCCACTGGACCGGAGAGATTGCCAAG 593
181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
594 TATGAGCTCTTTGGCCACGCTGTGTACAGAAATGGTGCCTCAGTGGAGGCCCATGAAC 653
201 IleSerIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
654 ATCTCCATCATCGTACCGACCAAGATGACCAAGCCCAAGTTTACCCAGGACACCTTC 713
221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
714 CGAGGGAGTGTCTTAGAGGGAGTCCCTACCAAGGTACTTCTGTGATGAGGTGACAGCCACA 773
241 AspGluAspAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln 260
774 GATGAGGATGATGCCATCTACACCTACCAATGGGGTGGTGTCTTACTCCATCCATAGCCAA 833
261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
834 GAACCAAGAGCCACACAGCACTCATGTTCCACAATTCACCGGAGCACAGGCACCATCAGC 893
281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
894 GTCATCTCCAGTGGCTTGGACCGGGAAGATCCCTGAGTACACATGACCATCCAGGCC 953
301 ThrAspMetAspGlyAspGlySerThrThrAlaValAlaValValGluIleLeuAsp 320
954 ACAGACATGATGGGACGGCTCCACCACCGCAGTGCGAGTAGTGGAGATCTTGAT 1013
321 AlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
1014 GCCAATGACAATGCTCCATGTTTGACCCCGAAGTACAGGCCCATGTGCGCTGAGAAT 1073
341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
1074 GCAGTGGGCCATGAGGTGCAGAGGCTGACGGTCACTGATCTGCACGCCCCCACTCACCA 1133
361 AlaTyrArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
1134 GCCTGGCGTCCCACTACCTTATCATGGCGGTGACGACGGGACCATTTTACCATCACCC 1193
381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
1194 ACCACCTCTGAGACACACAGGCACTCTGCACAAACAGAGAGGTTTGATTTGAGGCC 1253
401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
1254 AAAAAACAGACACACCTGTGACGTTGAAGTACCAACGAGGCCCTTTTGTGCTGAAGCTC 1313
421 ProThrSerThrAlaThrIleValHisValGluAspValAsnGluAlaProValPhe 440
1314 CCAACCTCCACAGCCCATAGTGGTCCAGGTGGAGATGTGAATGAGGACCTGTGTTT 1373
441 ValProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
1374 GTCCACCTCCAAAGTGTGTGAGTCCAGAGGGCATCCCCACTCGGGAGCCTGTGTCT 1433
461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
1434 GTTACACTGCAGAACCCCTGACAAAGGAGAATCAAAAGATACGCTACCGCATCTCTGAGA 1493

481 AspProAlaGlyTyrIleLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
1494 GACCCAGCAGGGTGGCTAGCCATGGACCCAGACAGTGGSCAGGTACAGACTGTGGGCACC 1553
501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
1554 CTCGACCGTGCAGATGAGCAGTGTGTGAGGAACAACATCTATGAAGTCAATGGCTTGGCC 1613
521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuLeuAsp 540
1614 ATGGCAATGGAAGCCCTCCACCACTGGCAACGGAAACCTTCTGCTTAACACTGATTGAT 1673
541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
1674 GTCACAGCATGGCCCGGCTGCTGAGCCCGCTCAGATCACCATCTGCAACCAAGCCCT 1733
561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
1734 GTGGCCACGCTGCTGAACATCAGCAGCAAGGACCTGTCTCCCAACACCTCCCTTCCAG 1793
581 AlaGlnLeuThrAspAspSerAspIleTyrThrAlaGluValAsnGluGlyAsp 600
1794 GCCCAGCTCACAGATGACTCAGACATCTACTGACGGCAGAGGTCAACGAGGAAGGTGAC 1853
601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
1854 ACAGTGGTCTTGTCCCTGAAGAGTCTCTGAAGCAGGATACATATGACGTGCACCTTCT 1913
621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
1914 CTGTCTGACCATGGCAACAAAGAGAGCTGACGGTGTATCAGGGCCACTGTGTGCACATGC 1973
641 HisGlyHisValGluThrCysProGlyProThrLysGlyGlyPheIleLeuProValLeu 660
1974 CATGGCCATGTCCAAACCTGCCCTGGACCTGGAAAGAGGTTTCTCTCCCTGTGTGTG 2033
661 GlyAlaValLeuAlaLeuPheLeuLeuValLeuLeuLeuValLeuValArgLysLys 680
2034 GGGGCTGTCTGGCTCTGCTGTCTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2093
681 ArgLysIleLysGluProLeuLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
2094 CGGAAGATCAAGAGGCCCTCTCTCTCCAGAAGATGACCCCGTGACACCGTCTCTTCTAC 2153
701 TyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
2154 TATGGCGAAGAGGGGGTGGCGAAGAGCAGGACCTATGACATCACCAGCTCCACCGA 2213
721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
2214 GGTCTGGAGGCCAGGCCGAGGTGTCTCTCGCAATGACGTGGCACCACCAACCATCATCCCG 2273
741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
2274 ACACCATGTACCTCTAGGCCAGCCAAACCCAGATGAAATCGGCACCTTTATTAATTGAG 2333
761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
2334 AACCTGAAGCGGCTTAACACAGACCCACAGCCCGCCCTACGACACCTCTTGTGTGTTC 2393
781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
2394 GACTATGAGGGGAGCGGCTCCGACCGCGCTCCCTGAGCTCCCTCACCCTCTCCGCGCTCC 2453
801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTyrGlySerArgPheLysLysLeuAla 820
2454 GACCAAGACCAAGATTACGATTATCTGAACGAGTGGGGCAGCCCGCTTCAAGAGCTGGCA 2513
821 AspMetTyrGlyGlyGlyAspAsp 829
2514 GACATGTACGTGGCGGGGAGGACGAC 2540


```
Db      ||||| GTGGCCACGTGTGAACATCAGGACAGACCTGTCTCCCCACACCTCCCTTTCCAG 2355
Qy      581 AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGlyAsp 600
Db      2356 GCCCAGCTCACAGATGACTCAGACATCTACTGGACGGCAGAGGTCAACGAGGAAGGTGAC 2415
Qy      601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
Db      2416 ACAGTGTCTTGTCCCTGAAGAGTTCCTGAGCAGGATACATATGACGTGCACCTTTCT 2475
Qy      621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
Db      2476 CTGTCTGACCATGCAACAAGAGCAGCTGACGGTGCATCAGGGCCACTGTGTCCGACTGC 2535
Qy      641 HisGlyHisValGluThrCysProGlyProTyrLysGlyGlyPheIleLeuProValLeu 660
Db      2536 CATGGCCATGTGNAACCTGCCCTGGACCTCGAAGGGAGGTTTCATCCCTCCCTGTGCTG 2595
Qy      661 GlyAlaValLeuAlaLeuPheLeuLeuLeuValLeuLeuLeuValArgLysLys 680
Db      2596 GGGGCTGTCTGTGCTGTCTGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2655
Qy      681 ArgLysIleLysGluProLeuLeuLeuProGluAspThrArgAspAsnValPheTyr 700
Db      2656 CGGAAGATCAAGAGAGCCCTCTCTACTCCCAAGATGACACCCGTGACACACGTCCTTCTAC 2715
Qy      701 TyrGlyGluGluGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
Db      2716 TATGGCGAAGGGGGTGGCGAAGAGGACGAGGACTATGACATCACCAGCTCCACCGA 2775
Qy      721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
Db      2776 GGTCTGAGGCCAGGGCGGAGGTGGTCTCCGCAATGACGTGGCACCACCATCATCCCG 2835
Qy      741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
Db      2836 ACACCCATGTACCGTCTAGCCAGCCACCCAGATGAAATCGGCACCTTTATATTGAG 2895
Qy      761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
Db      2896 AACCTGAAGCGCGCTAACACAGACCCGACAGCCCGCCCTACGACACCCCTCTGGTGTTC 2955
Qy      781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerAlaSer 800
Db      2956 GACTATGAGGGCAGCGGCTCCGACGCGCGCTCCCTGAGCTCCCTCACCTCCCTCCGCTCC 3015
Qy      801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGlySerArgPheLysLysLeuAla 820
Db      3016 GACCAAGACCAAGATTACGATTATCTGACGAGTGGGGCAGCGCTTCAAGAAGCTGGCA 3075
Qy      821 AspMetTyrGlyGlyGluAspAsp 829
Db      3076 GACATGTACGGTGGCGGGAGGAGCAG 3102
```

Search completed: September 23, 2004, 10:29:38
Job time : 1895 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 22:18:25 ; Search time 46 Seconds
(without alignments)
1733.539 Million cell updates/sec

Title: US-09-916-849A-1
Perfect score: 829
Sequence: 1 MGLPRGLASLLQLQVCWLQ.....NEWGSRFKLADWYGGEDD 829

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 0
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR 78:**
1: Pir1:**
2: Pir2:**
3: Pir3:**
4: Pir4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	829	100.0	829	1 IJHUCP	cadherin 3 precurs
2	61	7.4	491	1 IJBOCP	P-cadherin - bovin
3	51	6.2	822	1 IJMSCP	P-cadherin precurs
4	17	2.1	732	1 IJHCB	B-cadherin precurs
5	17	2.1	887	1 IJHCL	E-cadherin precurs
6	16	1.9	895	1 IJLCP	Ep-cadherin precurs
7	16	1.9	905	2 S43064	cadherin - African
8	15	1.8	871	2 S47518	cadherin - African
9	15	1.8	882	1 IJHUCE	cadherin 1 precurs
10	15	1.8	884	1 IJMSCE	E-cadherin precurs
11	15	1.8	894	2 S34438	uvomorulin - mouse
12	13	1.6	913	1 A47543	R-cadherin precurs
13	13	1.6	913	1 IJHCR	R-cadherin precurs
14	13	1.6	916	2 C38992	cadherin 4 precurs
15	12	1.4	783	2 I50116	N-cadherin precurs
16	12	1.4	877	1 IJBOCN	N-cadherin precurs
17	12	1.4	905	1 IJLCL1	N-cadherin 1 precu
18	12	1.4	906	1 IJHUCN	cadherin 2 precurs
19	12	1.4	906	1 IJMSCN	N-cadherin precurs
20	12	1.4	906	1 IJLCL2	N-cadherin 2 precu
21	12	1.4	912	1 I52701	N-cadherin precurs
22	11	1.3	789	2 I52701	K-cadherin - rat
23	11	1.3	790	2 I37016	cadherin-6 - human
24	10	1.2	353	2 T07998	cytochrome c-type
25	10	1.2	353	2 T08001	cytochrome c-type
26	10	1.2	790	2 I50178	cadherin-6B - chic
27	9	1.1	481	2 C95820	hypothetical membr
28	9	1.1	3034	2 T14119	seven-pass transme
29	8	1.0	40	2 D83378	hypothetical prote

30 8 1.0 103 2 A26736 transformation-ind
31 8 1.0 132 2 H75548 hypothetical prote
32 8 1.0 139 2 F85565 protein R10E11.9 [
33 8 1.0 157 2 A86413 probable ribosomal
34 8 1.0 182 2 T46396 hypothetical prote
35 8 1.0 243 2 D64188 arginine transport
36 8 1.0 279 2 G83196 glycerol uptake fa
37 8 1.0 293 2 T32892 hypothetical prote
38 8 1.0 340 2 T32891 hypothetical prote
39 8 1.0 364 2 F70137 cell division prot
40 8 1.0 379 2 T11375 ubiquinol-cytochro
41 8 1.0 384 2 B83443 probable acyl-CoA
42 8 1.0 393 2 T01371 hypothetical prote
43 8 1.0 398 2 S77230 hypothetical prote
44 8 1.0 463 2 A70469 probable permease
45 8 1.0 518 2 S55948 hypothetical prote

ALIGNMENTS

RESULT 1

IJHUCP
cadherin 3 precursor - human
N:Alternate names: P-cadherin; placental cadherin
C:Species: Homo sapiens (man)
C:Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C:Accession: A33659
R:Shimoyama, Y.; Yoshida, T.; Terada, M.; Shimamoto, Y.; Abe, O.; Hirohashi, S.
J. Cell Biol. 109, 1787-1794, 1989
A:Title: Molecular cloning of a human Ca(2+)-dependent cell-cell adhesion molecule homolo
A:Reference number: A33659; MUID:90009051; PMID:2793940
A:Accession: A33659
A:Molecule type: mRNA
A:Residues: 1-829 <SHI>
A:Cross-references: GB:X63629; NID:G35322; PIDN:CAA45177.1; PID:G35323
C:Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought to
C:Genetics:
A:Gene: GDB:CDH3
A:Cross-references: GDB:I32860; OMIM:114021
A:Map position: 16q24.1-16qter
C:Superfamily: cadherin; cadherin repeat homology
C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane prot
F:1-46/Domain: signal sequence #status predicted <SIG>
F:127-107/Domain: propeptide #status predicted <PRO>
F:108-829/Product: P-cadherin #status predicted <PRAT>
F:108-650/Domain: extracellular #status predicted <EXT>
F:110-215/Domain: cadherin repeat homology <CR1>
F:185-190/Region: cadherin binding #status predicted
F:218-328/Domain: cadherin repeat homology <CR2>
F:331-440/Domain: cadherin repeat homology <CR3>
F:441-548/Domain: cadherin repeat homology <CR4>
F:549-652/Domain: cadherin repeat homology <CR5>
F:653-677/Domain: transmembrane #status predicted <TMM>
F:678-829/Domain: intracellular #status predicted <INT>
F:785-800/Region: serine-rich
F:200,566/Binding site: carbohydrate (Asn) #status predicted

Query Match 100.0%; Score 829; DB 1; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGLPRGLASLLQLQVCWLQCAASEPCRAVFREAEVTLAAGAEQEPQALGKVFMCPCG 60
Db 1 MGLPRGLASLLQLQVCWLQCAASEPCRAVFREAEVTLAAGAEQEPQALGKVFMCPCG 60
Qy 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKILRRHKRDVWVAPISVPENG 120
Db 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKILRRHKRDVWVAPISVPENG 120
Qy 121 KGPFPQRLNQLKSNKDRDRTKIFYSITGPGADSPPEGVFAVEKETGWLILNKLPLDREIAK 180
Db 121 KGPFPQRLNQLKSNKDRDRTKIFYSITGPGADSPPEGVFAVEKETGWLILNKLPLDREIAK 180

```

Query Match          7.4%; Score 61; DB 1; Length 491;
Best Local Similarity 100.0%; Pred. No. 1.3e-51;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 676 LVRRKRIKIPLLPDDTRDNVFFYGGEGGEDQDYDITQLHRGLEARPEVVLNDVA 735
DB 338 LVRRKRIKIPLLPDDTRDNVFFYGGEGGEDQDYDITQLHRGLEARPEVVLNDVA 397
QY 736 P 736
DB 398 P 398

RESULT 3
LJMSCP
P-cadherin precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 07-May-1999
C:Accession: S03163; S34458
R:Nose, A.; Nagafuchi, A.; Takeichi, M.
EMBO J. 6, 3655-3661, 1987
A:Title: Isolation of placental cadherin cDNA: identification of a novel gene family of
A:Reference number: S03163; MUID:88111554; PMID:3428270
A:Accession: S03163
A:Molecule type: mRNA
A:Residues: 1-822 <NOS>
A:Cross-references: EMBL:X06340
R:Faraldo, M.L.M.; Cano, A.
J. Mol. Biol. 231, 935-941, 1993
A:Title: The 5' flanking sequences of the mouse P-cadherin gene. Homologies to 5' sequen
A:Reference number: S34458; MUID:93294853; PMID:8515462
A:Accession: S34458
A:Molecule type: DNA
A:Residues: 1-55 <FAR>
A:Cross-references: EMBL:X68057
C:Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought to
C:Genetics:
A:Introns: 16/3
A:Superfamily: cadherin; cadherin repeat homology
C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; placenta; transmem
F:1-27/Domain: signal sequence #status predicted <SIG>
F:28-99/Domain: propeptide #status predicted <PRO>
F:100-822/Product: P-cadherin #status predicted <MAT>
F:100-645/Domain: extracellular #status predicted <EXT>
F:102-207/Domain: cadherin repeat homology <CR1>
F:210-320/Domain: cadherin repeat homology <CR2>
F:323-432/Domain: cadherin repeat homology <CR3>
F:433-540/Domain: cadherin repeat homology <CR4>
F:541-645/Domain: cadherin repeat homology <CR5>
F:646-676/Domain: transmembrane #status predicted <TMM>
F:671-822/Domain: intracellular #status predicted <INT>
F:778-793/Region: serine-rich
F:192,558/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match          6.2%; Score 51; DB 1; Length 822;
Best Local Similarity 100.0%; Pred. No. 1.6e-41;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 684 KEPLLPDDTRDNVFFYGGEGGEDQDYDITQLHRGLEARPEVVLNDV 734
DB 677 KEPLLPDDTRDNVFFYGGEGGEDQDYDITQLHRGLEARPEVVLNDV 727

RESULT 4
LJCHCB
B-cadherin precursor - chicken (fragment)
N:Alternate names: K-CAM protein
C:Species: Gallus gallus (chicken)
C:Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C:Accession: A41634; A38715; S16160
R:Sorkin, B.C.; Gallin, W.J.; Edelman, G.M.; Cunningham, B.A.
Proc. Natl. Acad. Sci. U.S.A. 88, 11545-11549, 1991
A:Title: Genes for two calcium-dependent cell adhesion molecules have similar structures

```


A:Reference number: A41634; MUID:92107987; PMID:1763068
A:Accession: A41634
A:Molecule type: DNA
A:Residues: 1-732 <SOR>
A:Cross-references: GB:M81894; NID:9212226; PIDN:AAA48925.1; PID:G212227
R:Napolitano, E.W.; Venstrom, K.; Wheeler, E.F.; Reichardt, L.F.
J. Cell Biol. 113, 893-905, 1991
A:Title: Molecular cloning and characterization of B-cadherin, a novel chick cadherin.
A:Reference number: A38715; MUID:91225083; PMID:2026653
A:Accession: A38715
A:Molecule type: mRNA
A:Residues: 7-413, 'V', 415-732 <NAP>
A:Cross-references: GB:X58518; NID:G631113; PIDN:CAA41408.1; PID:G631114
C:Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought to be involved in a variety of other cellular processes.
C:Keywords: Cadherins mediate calcium-dependent intercellular adhesion, and are thought to be involved in a variety of other cellular processes.
A:Gene: K-CAM
A:Introns: 29/3; 81/3; 130/1; 188/3; 231/3; 292/3; 375/2; 423/1; 498/1; 571/1; 614/3; 668/1
C:Superfamily: cadherin; cadherin repeat homology
C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane protein
F:1-6/Domain: extracellular #status predicted <PRO>
F:6-554/Domain: extracellular #status predicted <EXT>
F:7-732/Product: B-cadherin #status predicted <MAT>
F:9-114/Domain: cadherin repeat homology <CR1>
F:84-89/Region: cadherin binding #status predicted
F:117-227/Domain: cadherin repeat homology <CR2>
F:230-339/Domain: cadherin repeat homology <CR3>
F:340-447/Domain: cadherin repeat homology <CR4>
F:448-552/Domain: cadherin repeat homology <CR5>
F:555-580/Domain: transmembrane #status predicted <TM>
F:581-732/Domain: intracellular #status predicted <INT>
F:689-702/Region: serine-rich
F:137,410/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 2.1%; Score 17; DB 1; Length 732;
Best Local Similarity 100.0%; Pred. No. 6e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 797 SSASDQDDYDYLNEWG 813
DB 699 SSASDQDDYDYLNEWG 715

RESULT 5
IUCHLC
E-cadherin precursor, hepatic - chicken
N:Alternate names: L-CAM; liver cell adhesion molecule
C:Species: Gallus gallus (chicken)
C:Date: 30-Jun-1993 #sequence revision 30-Jun-1993 #text_change 25-Oct-1996
C:Accession: A30201; A29866; E29866
R:Sorkin, B.C.; Hemperly, J.J.; Edelman, G.M.; Cunningham, B.A.
Proc. Natl. Acad. Sci. U.S.A. 85, 7617-7621, 1988
A:Title: Structure of the gene for the liver cell adhesion molecule, L-CAM.
A:Reference number: A30201; MUID:89017248; PMID:3174655
A:Accession: A30201
A:Molecule type: mRNA
A:Residues: 1-81 <SOR>
A:Cross-references: EMBL:J04074
R:Gallin, W.J.; Sorkin, B.C.; Edelman, G.M.; Cunningham, B.A.
Proc. Natl. Acad. Sci. U.S.A. 84, 2808-2812, 1987
A:Title: Sequence analysis of a cDNA clone encoding the liver cell adhesion molecule, L-CAM.
A:Reference number: A29866; MUID:87204217; PMID:3472238
A:Accession: A29866
A:Molecule type: mRNA
A:Residues: 51-887 <GAL>
A:Cross-references: EMBL:M16260
A:Accession: B29866
A:Molecule type: protein
A:Residues: 161-172; 323-336; 386-407; 533-551 <GA2>
C:Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought to be involved in a variety of other cellular processes.
C:Genetics: 23/3; 62/1; 138/1 183/3; 235/3; 284/1; 342/3; 385/3; 447/3; 529/2; 577/1; 652/1
A:Introns: 23/3; 62/1; 138/1 183/3; 235/3; 284/1; 342/3; 385/3; 447/3; 529/2; 577/1; 652/1
C:Superfamily: cadherin; cadherin repeat homology
C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; liver; transmembrane protein

F:1-26/Domain: signal sequence #status predicted <SIG>
F:27-160/Domain: propeptide #status predicted <PRO>
F:161-887/Product: E-cadherin, hepatic #status experimental <MAT>
F:161-704/Domain: extracellular #status predicted <EXT>
F:163-268/Domain: cadherin repeat homology <CR1>
F:238-243/Region: cadherin binding #status predicted
F:271-381/Domain: cadherin repeat homology <CR2>
F:384-493/Domain: cadherin repeat homology <CR3>
F:494-601/Domain: cadherin repeat homology <CR4>
F:602-704/Domain: cadherin repeat homology <CR5>
F:705-735/Domain: transmembrane #status predicted <TM>
F:736-887/Domain: intracellular #status predicted <INT>
F:844-857/Region: serine-rich
F:291,346,564,643/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 2.1%; Score 17; DB 1; Length 887;
Best Local Similarity 100.0%; Pred. No. 7.1e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 797 SSASDQDDYDYLNEWG 813
DB 854 SSASDQDDYDYLNEWG 870

RESULT 6
IUXLCP
EP-cadherin precursor - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 30-Jun-1993 #sequence revision 30-Jun-1993 #text_change 22-Jun-1999
C:Accession: B43785; A60128
R:Ginsberg, D.; Desimone, D.; Geiger, B.
Development 111, 315-325, 1991
A:Title: Expression of a novel cadherin (EP-cadherin) in unfertilized eggs and early Xenopus laevis.
A:Reference number: A43785; MUID:91372132; PMID:1893866
A:Accession: B43785
A:Molecule type: mRNA
A:Residues: 1-895 <GIN>
A:Cross-references: GB:X63720; NID:G64681; PIDN:CAA45252.1; PID:G64682
A:Note: It is uncertain whether Met-1 or Met-16 is the initiator
R:Angres, B.; Mueller, A.H.J.; Kellermann, J.; Hausen, P.
Development 111, 829-844, 1991
A:Title: Differential expression of two cadherins in Xenopus laevis.
A:Reference number: A60128; MUID:91347911; PMID:1879345
A:Accession: A60128
A:Molecule type: protein
A:Residues: 171-177, 'I', 179-183, 'K', 185-189, 'XI' <ANG>
A:Note: the material sequenced may have contained U-cadherin as well as E-cadherin
C:Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought to be involved in a variety of other cellular processes.
C:Superfamily: cadherin; cadherin repeat homology
C:Keywords: calcium binding; cell adhesion; duplication; embryo; glycoprotein; transmembrane protein
F:1-43/Domain: signal sequence #status predicted <SIG>
F:44-170/Domain: propeptide #status predicted <PRO>
F:171-895/Product: EP-cadherin #status predicted <MAT>
F:171-718/Domain: extracellular #status predicted <EXT>
F:173-278/Domain: cadherin repeat homology <CR1>
F:248-253/Region: cadherin binding #status predicted
F:281-391/Domain: cadherin repeat homology <CR2>
F:394-502/Domain: cadherin repeat homology <CR3>
F:503-610/Domain: cadherin repeat homology <CR4>
F:611-714/Domain: cadherin repeat homology <CR5>
F:719-743/Domain: transmembrane #status predicted <TM>
F:744-895/Domain: intracellular #status predicted <INT>
F:852-865/Region: serine-rich
F:440,696/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 1.9%; Score 16; DB 1; Length 895;
Best Local Similarity 100.0%; Pred. No. 7e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 699 FYGEGGGGEBDDYD 714
DB 765 FYGEGGGGEBDDYD 780

RESULT 7

S43064
 Cadenherin - African clawed frog
 C/Species: Xenopus laevis (African clawed frog)
 C/Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 20-Aug-1999
 C/Accession: S43064; S43065
 R/Kueh, M.
 submitted to the EMBL Data Library, March 1994
 A/Reference number: S43064
 A/Accession: S43064
 A/Molecule type: mRNA
 A/Residues: 1-905 <KUE>
 A/Cross-references: EMBL:X78546; NID:G468816; PIDN:CAA55292.1; PID:G468817
 R/Herzberg, F.; Wildermuth, V.; Wedlich, D.
 Mech. Dev. 35, 33-42, 1991
 A/Title: Expression of XBCad, a novel cadherin, during oogenesis and early development
 A/Reference number: S43065; MUID:92062581; PMID:1840622
 A/Accession: S43065
 A/Molecule type: mRNA
 A/Residues: 'NSA', 462-697, 'Q', 699-807, 'A', 809-840, 'N', 842-877, 'N', 885-902, 'E'
 A/Cross-references: EMBL:X78546
 C/Superfamily: cadherin; cadherin repeat homology
 C/Keywords: calcium binding; cell adhesion; glycoprotein; transmembrane protein
 F/181-286/Domain: cadherin repeat homology <CR1>
 F/289-399/Domain: cadherin repeat homology <CR2>
 F/402-510/Domain: cadherin repeat homology <CR3>
 F/511-618/Domain: cadherin repeat homology <CR4>
 F/619-722/Domain: cadherin repeat homology <CR5>

Query Match 1.9%; Score 16; DB 2; Length 905;
 Best Local Similarity 100.0%; Pred. No. 7.1e-07;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 699 FYTGEEGGEGEDQDYD 714
 |||||

DB 773 FYTGEEGGEGEDQDYD 788
 |||||

RESULT 8

S47518
 Cadenherin - African clawed frog
 C/Species: Xenopus laevis (African clawed frog)
 C/Date: 01-Feb-1995 #sequence_revision 10-Nov-1995 #text_change 17-Mar-1999
 C/Accession: S47518
 R/Tsui, O.; Fujii, G.; Tashiro, K.; Shiohawa, K.
 Biochim. Biophys. Acta 1219, 121-128, 1994
 A/Title: Molecular cloning of cDNA for XTCAD-1, a novel Xenopus cadherin, and its expression
 A/Reference number: S47518; MUID:94368839; PMID:8086449
 A/Accession: S47518
 A/Molecule type: mRNA
 A/Residues: 1-871 <TOO>
 C/Superfamily: cadherin; cadherin repeat homology
 F/151-256/Domain: cadherin repeat homology <CR>

Query Match 1.8%; Score 15; DB 2; Length 871;
 Best Local Similarity 100.0%; Pred. No. 6.7e-06;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 744 YRPRPNDPDEIGNFI 758
 |||||

DB 787 YRPRPNDPDEIGNFI 801
 |||||

RESULT 9

LUHUCE
 Cadenherin 1 precursor [invalidated] - human
 N/Alternate names: ARC-1; Cell CAM 120/80; E-cadherin; epithelial cadherin; L-CAM; uvomor-
 C/Species: Homo sapiens (man)
 C/Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 08-Dec-2000
 C/Accession: S37654; S31430; S05475; S31460; S06716; A57171; JC2230; I52294; I52704; S25
 R/Bussemaekers, M.J.G.; van Bokhoven, A.; Mees, S.G.M.; Kemler, R.; Schalken, J.A.
 Mol. Biol. Rep. 17, 123-128, 1993

A/Title: Molecular cloning and characterization of the human E-cadherin cDNA.
 A/Reference number: S37654; MUID:93211394; PMID:8459805
 A/Accession: S37654
 A/Molecule type: mRNA

A/Residues: 1-882 <BUS>
 A/Cross-references: EMBL:Z13009; NID:G31072; PIDN:CAA78353.1; PID:G31073
 R/Kemler, W.; Warda, A.; Oda, T.; Hirohashi, S.; Kemler, R.; Birchmeier, W.

submitted to the EMBL Data Library, December 1992
 A/Description: Sequence of human E-cadherin cDNA.
 A/Reference number: S31430
 A/Accession: S31430
 A/Molecule type: mRNA

A/Residues: 1-542, 'F', 544-882 <KEL>
 A/Cross-references: EMBL:Z18923; NID:G31074; PIDN:CAA79356.1; PID:G31075
 R/Mansouri, A.; Spurr, N.; Goodfellow, P.N.; Kemler, R.

Differentiation 38, 67-71, 1988
 A/Title: Characterization and chromosomal localization of the gene encoding the human cel
 A/Reference number: S05475; MUID:89031725; PMID:3263290
 A/Accession: S05475
 A/Molecule type: mRNA

A/Residues: 157-311 <MAN>
 A/Cross-references: EMBL:X12790
 A/Note: nucleotide sequence is not complete

R/Frixen, U.H.
 submitted to the EMBL Data Library, March 1990
 A/Reference number: S31460
 A/Accession: S31460

A/Molecule type: mRNA
 A/Residues: 265-392 <FRI>
 A/Cross-references: EMBL:X52279; NID:G28821; PIDN:CAA36522.1; PID:G28822

R/Wheelerlock, M.J.; Buck, C.A.; Bechtol, K.B.; Damsky, C.H.
 J. Cell. Biochem. 34, 187-202, 1987

A/Title: Soluble 80-kd fragment of cell-CAM 120/80 disrupts cell-cell adhesion.
 A/Reference number: S06716; MUID:87280410; PMID:3611200
 A/Accession: S06716

A/Molecule type: protein
 A/Residues: 'XQ', 157-162, 'V', 164-179 <WHE>
 R/Berx, G.; Staes, K.; van Hengel, J.; Molemans, F.; Bussemakers, M.J.G.; van Bokhoven, J

Genomics 26, 281-289, 1995
 A/Title: Cloning and characterization of the human invasion suppressor gene E-cadherin (C
 A/Reference number: A57171; MUID:95324920; PMID:7601454

A/Accession: A57171
 A/Status: nucleic acid sequence not shown; not compared with conceptual translation

A/Molecule type: DNA
 A/Residues: 1-30, 32-882 <BER>
 A/Cross-references: GB:L34784

R/Rimm, D.L.; Morrow, J.S.
 Biochem. Biophys. Res. Commun. 200, 1754-1761, 1994

A/Title: Molecular cloning of human E-cadherin suggests a novel subdivision of the cadhe
 A/Reference number: JC2230; MUID:94242050; PMID:8185635
 A/Accession: JC2230

A/Molecule type: mRNA
 A/Residues: 1-9, 'G', 11-15, 'RSPGSGRSPPCLTRHLVHGAPAPKPR', 52-67, 'I', 69, 'ITPIP', 76-94,
 A/Cross-references: GB:L08599; NID:G340184; PIDN:AAA61259.1; PID:G340185

A/Note: the majority of differences between this and other reports represent apparent fra
 A/Note: the authors translated the codon CCG for residue 868 as Arg

R/Bussemaekers, M.J.G.; Groidi, L.A.; van Bokhoven, A.; Schalken, J.A.
 Biochem. Biophys. Res. Commun. 203, 1284-1290, 1994

A/Title: Transcriptional regulation of the human E-cadherin gene in human prostate cancer
 A/Reference number: I52294; MUID:94380041; PMID:8093045
 A/Accession: I52294

A/Status: translation not shown; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-16 <RES>

A/Cross-references: GB:L34545; NID:G509604; PIDN:AAA21764.1; PID:G509605
 R/Becker, K.F.; Atkinson, M.J.; Reich, U.; Nekarda, H.; Stewert, J.R.; Hoflei

Cancer Res. 54, 3845-3852, 1994
 A/Title: E-cadherin gene mutations provide clues to diffuse type gastric carcinomas.

A/Reference number: I52704; MUID:94306394; PMID:8033105
 A/Accession: I52704
 A/Status: translated from GB/EMBL/DBJ

A/Molecule type: DNA
 A/Residues: 337-476 <RE2>

A;Cross-references: GB:S72492; NID:G632756; PIDN:AAI4108.1; PID:G4261808
C;Comment: Cadherins mediate calcium-dependent intercellular adhesion and are thought to
C;Genetics:
A;Gene: GDB:CDH1; UVO
A;Cross-references: GDB:120484; OMIM:192090
A;Map position: 16q22.1-16q22.1
A;Introns: 379/3; 440/3
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane protein
F;1-27/Domain: signal sequence #status predicted <SIG>
F;128-156/Domain: propeptide #status predicted <PRO>
F;155-882/Product: E-cadherin #status experimental <WAT>
F;155-697/Domain: extracellular #status predicted <EXT>
F;157-262/Domain: cadherin repeat homology <CR1>
F;232-237/Region: cadherin binding #status predicted
F;265-375/Domain: cadherin repeat homology <CR2>
F;378-486/Domain: cadherin repeat homology <CR3>
F;487-595/Domain: cadherin repeat homology <CR4>
F;596-700/Domain: cadherin repeat homology <CR5>
F;698-731/Domain: transmembrane #status predicted <TM>
F;732-882/Domain: intracellular #status predicted <INT>
F;840-853/Region: serine-rich
F;376,558,570,622,637,849/Binding site: carbohydrate (Asn) (covalent) #status predicted
Query Match 1.8%; Score 15; DB 1; Length 882;
Best Local Similarity 100.0%; Pred. No. 6.7e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 815 RPKKADMYGGEDD 829
DB 868 RPKKADMYGGEDD 882
RESULT 10
IJMSCE
E-cadherin precursor, epithelial - mouse
N;Alternate names: uvomorulin
C;Species: Mus musculus (house mouse)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C;Accession: S04528; S03160; I49565; S48735
R;Nagafuchi, A.; Shirayoshi, Y.; Okazaki, K.; Yasuda, K.; Takeichi, M.
Nature 329, 341-343, 1987
A;Title: Transformation of cell adhesion properties by exogenously introduced E-cadherin
A;Reference number: S04528; MUID:87315445; PMID:3498123
A;Accession: S04528
A;Molecule type: mRNA
A;Residues: 1-412, 'V', 414-884 <NAG>
A;Cross-references: EMBL:X06115
R;Ringwald, M.; Schuh, R.; Vestweber, D.; Bistetter, H.; Lottspeich, F.; Engel, J.; Doel
EMBO J. 6, 3647-3653, 1987
A;Title: The structure of cell adhesion molecule uvomorulin. Insights into the molecular
A;Reference number: S03160; MUID:88111553; PMID:3501370
A;Accession: S03160
A;Molecule type: mRNA
A;Residues: 157-884 <RIN>
A;Cross-references: EMBL:X06339
A;Note: part of this sequence, including the amino end of the mature protein, was confir
R;Behrens, J.; Loewrick, O.; Klein-Hitpass, L.; Birchmeier, W.
Proc. Natl. Acad. Sci. U.S.A. 88, 11495-11499, 1991
A;Title: The E-cadherin promoter: Functional analysis of a G-C-rich region and an epithe
A;Reference number: I49565; MUID:92107977; PMID:1763063
A;Accession: I49565
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-15 <RES>
A;Cross-references: GB:M81449; NID:G192325; PIDN:AAA37352.1; PID:G192326
R;Tong, K.I.; Yau, P.; Overduin, M.; Badby, S.; Porumb, T.; Takeichi, M.; Ikura, M.
FEBS Lett. 352, 318-322, 1994
A;Title: Purification and spectroscopic characterization of a recombinant amino-terminal
A;Reference number: S48735; MUID:95010732; PMID:7925993
A;Accession: S48735
A;Status: preliminary
A;Molecule type: protein

A;Residues: 156-300 <TON>
C;Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought to
C;Genetics:
A;Gene: E-cadherin
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane protein
F;1-27/Domain: signal sequence #status predicted <SIG>
F;128-156/Domain: propeptide #status predicted <PRO>
F;157-884/Product: E-cadherin, epithelial #status experimental <WAT>
F;157-699/Domain: extracellular #status predicted <EXT>
F;159-264/Domain: cadherin repeat homology <CR1>
F;234-237/Region: cadherin binding #status predicted
F;267-377/Domain: cadherin repeat homology <CR2>
F;380-488/Domain: cadherin repeat homology <CR3>
F;489-597/Domain: cadherin repeat homology <CR4>
F;598-702/Domain: cadherin repeat homology <CR5>
F;702-733/Domain: transmembrane #status predicted <TM>
F;734-884/Domain: intracellular #status predicted <INT>
F;842-855/Region: serine-rich
F;560,639/Binding site: carbohydrate (Asn) (covalent) #status predicted
Query Match 1.8%; Score 15; DB 1; Length 884;
Best Local Similarity 100.0%; Pred. No. 6.8e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 744 YRPRANPDEIGNFI 758
DB 759 YRPRANPDEIGNFI 813
RESULT 11
S34438
uvomorulin - mouse
C;Species: Mus musculus (house mouse)
C;Date: 22-Nov-1993 #sequence_revision 03-Nov-1995 #text_change 23-May-1997
C;Accession: S34438
R;Ringwald, M.; Baribault, H.; Schmidt, C.; Kemler, R.
Nucleic Acids Res. 19, 6533-6539, 1991
A;Title: The structure of the gene coding for the mouse cell adhesion molecule uvomorulin
A;Reference number: S34438; MUID:92093614; PMID:1754391
A;Accession: S34438
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-884 <RIN>
A;Cross-references: EMBL:X60975
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1992
C;Genetics:
A;Introns: 16/3; 57/1; 131/3; 179/3; 231/3; 280/1; 338/3; 381/3; 442/3; 524/2; 573/1; 64
C;Superfamily: cadherin; cadherin repeat homology
F;380-488/Domain: cadherin repeat homology <CR3>
Query Match 1.8%; Score 15; DB 2; Length 884;
Best Local Similarity 100.0%; Pred. No. 6.8e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 744 YRPRANPDEIGNFI 758
DB 799 YRPRANPDEIGNFI 813
RESULT 12
A47543
R-cadherin precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 03-May-1994 #sequence_revision 26-May-1994 #text_change 22-Jun-1999
C;Accession: A47543; I55501
R;Hutton, J.C.; Christofori, G.; Chi, W.Y.; Edman, U.; Guest, P.C.; Hanahan, D.; Kelly, F.
Mol. Endocrinol. 7, 1151-1160, 1993
A;Title: Molecular cloning of mouse pancreatic islet R-cadherin: differential expression
A;Reference number: A47543; MUID:94067164; PMID:8247017
A;Accession: A47543
A;Molecule type: mRNA
A;Residues: 1-913 <HUT>

A;Cross-references: EMBL:X69966; NID:G429111; PIDN:CAA49589.1; PID:G429112
 R;Matsunami, H.; Miyatani, S.; Inoue, T.; Copeland, N.; Gilbert, D.; Jenkins, N.; Takeid
 J. Cell Sci. 106, 401-409, 1993
 A;Title: Cell binding specificity of mouse R-cadherin and chromosomal mapping of the gen
 A;Reference number: I55501; MUID:94095672; PMID:8270638
 A;Accession: I55501
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1-913 <RES>
 A;Cross-references: GB:D14888; NID:G457658; PIDN:BA03605.1; PID:G457659
 C;Comment: Cadherins mediate calcium-dependent intercellular adhesion and are thought to
 C;Superfamily: cadherin; cadherin repeat homology
 C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane pro
 F;1-27/Domain: signal sequence #status predicted <SIG>
 F;28-166/Domain: propeptide #status predicted <PRO>
 F;167-913/Product: R-cadherin #status predicted <MAT>
 F;167-721/Domain: extracellular #status predicted <EXT>
 F;169-274/Domain: cadherin repeat homology <CR1>
 F;244-249/Region: cadherin binding #status predicted
 F;277-389/Domain: cadherin repeat homology <CR2>
 F;300-304/Domain: calcium binding #status predicted <CAB>
 F;392-504/Domain: cadherin repeat homology <CR3>
 F;507-612/Domain: cadherin repeat homology <CR4>
 F;613-721/Domain: cadherin repeat homology <CR5>
 F;722-753/Domain: transmembrane #status predicted <TM>
 F;754-913/Domain: intracellular #status predicted <INT>
 F;870-885/Region: serine-rich
 F;280,409,554,629,658,699/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 1.6%; Score 13; DB 1; Length 913;
 Best Local Similarity 100.0%; Pred. No. 0.00066;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 815 RFKKLADMYGGGE 827
 Db 899 RFKKLADMYGGGE 911

RESULT 13
 IJCHCR
 R-cadherin precursor - chicken
 C;Species: Gallus gallus (chicken)
 C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 16-Jun-2000
 A;Reference number: JH0424; MUID:91299341; PMID:1712604
 A;Accession: JH0424
 A;Molecule type: mRNA
 A;Residues: 1-913 <INU>
 A;Cross-references: GB:D14459; GB:D00849; NID:G222854; PIDN:BA03356.1; PID:G222855
 A;Experimental source: retina
 C;Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought b
 C;Superfamily: cadherin; cadherin repeat homology
 C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; retina; transmemb
 F;1-26/Domain: signal sequence #status predicted <SIG>
 F;27-166/Domain: propeptide #status predicted <PRO>
 F;167-913/Product: R-cadherin #status predicted <MAT>
 F;167-721/Domain: extracellular #status predicted <EXT>
 F;169-274/Domain: cadherin repeat homology <CR1>
 F;244-249/Region: cadherin binding #status predicted
 F;277-389/Domain: cadherin repeat homology <CR2>
 F;392-504/Domain: cadherin repeat homology <CR3>
 F;507-612/Domain: cadherin repeat homology <CR4>
 F;613-721/Domain: cadherin repeat homology <CR5>
 F;722-753/Domain: transmembrane #status predicted <TM>
 F;754-913/Domain: intracellular #status predicted <INT>
 F;870-885/Region: serine-rich
 F;280,409,554,629,658,699/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 1.6%; Score 13; DB 1; Length 913;
 Best Local Similarity 100.0%; Pred. No. 0.00066;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 815 RFKKLADMYGGGE 827
 Db 899 RFKKLADMYGGGE 911

RESULT 14
 C38992
 cadherin 4 precursor - human
 N;Alternate names: R-cadherin
 C;Species: Homo sapiens (man)
 C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 20-Aug-1999
 R;Suzuki, S.; Sano, K.; Tanihara, H.
 Cell Regul. 2, 261-270, 1991
 A;Title: Diversity of the cadherin family: evidence for eight new cadherins in nervous ti
 A;Reference number: S24305; MUID:91283540; PMID:2059658
 A;Accession: C38992
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1-916 <SUZ>
 A;Cross-references: GB:L34059; NID:G506409; PIDN:AAA35627.1; PID:G506410
 C;Genetics:
 A;Gene: GDB:CDH4
 A;Cross-references: GDB:622850
 A;Map position: 16q24.1-16qter
 C;Superfamily: cadherin; cadherin repeat homology
 C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane prot
 F;1-27/Domain: signal sequence #status predicted <SIG>
 F;28-166/Domain: propeptide #status predicted <PRO>
 F;167-913/Product: R-cadherin #status predicted <MAT>
 F;167-721/Domain: extracellular #status predicted <EXT>
 F;169-274/Domain: cadherin repeat homology <CR1>
 F;244-249/Region: cadherin binding #status predicted
 F;277-389/Domain: cadherin repeat homology <CR2>
 F;300-304/Domain: calcium binding #status predicted <CAB>
 F;392-504/Domain: cadherin repeat homology <CR3>
 F;507-612/Domain: cadherin repeat homology <CR4>
 F;613-721/Domain: cadherin repeat homology <CR5>
 F;722-753/Domain: transmembrane #status predicted <TM>
 F;754-913/Domain: intracellular #status predicted <INT>
 F;870-885/Region: serine-rich
 F;283,412,557,632,661,702/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 1.6%; Score 13; DB 2; Length 916;
 Best Local Similarity 100.0%; Pred. No. 0.00066;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 815 RFKKLADMYGGGE 827
 Db 902 RFKKLADMYGGGE 914

RESULT 15
 IS0116
 N-cadherin precursor - zebra fish
 C;Species: Brachydanio rerio (zebra fish)
 C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jan-2000
 C;Accession: IS0116
 R;Sitzur, S.; Kam, Z.; Geiger, B.
 Dev. Dyn. 201, 121-136, 1994
 A;Title: Structure and distribution of N-cadherin in developing zebrafish embryos: morpho
 A;Reference number: IS0116; MUID:95178741; PMID:7873785
 C;Accession: IS0116
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1-783 <BIT>
 A;Cross-references: EMBL:X67648; NID:9623209; PIDN:CAA47890.1; PID:G623210
 C;Superfamily: cadherin; cadherin repeat homology
 F;146-258/Domain: cadherin repeat homology <CDH>

Query Match 1.4%; Score 12; DB 2; Length 783;

Best Local Similarity 100.0%; Pred. No. 0.0056;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 703 EGGGEEDQDYD 714
Db 652 EGGGEEDQDYD 663

Search completed: September 21, 2004, 22:26:48
Job time : 47 secs.

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 22:17:15 ; Search time 26 Seconds
(without alignments)

1660.237 Million cell updates/sec

Title: US-09-916-849A-1

Perfect score: 829

Sequence: 1 MGIPRGPLASLLLLQVCWLQ.....NEWGSRFKLADMYGGEDD 829

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	829	100.0	829	1	CAD3_HUMAN
2	61	7.4	491	1	CAD3_BOVIN
3	51	6.2	822	1	CAD3_MOUSE
4	27	3.3	145	1	CAD3_PIG
5	17	2.1	732	1	CAD1_CHICK
6	17	2.1	887	1	CAD1_CHICK
7	16	1.9	880	1	CAD2_XENLA
8	16	1.9	884	1	CAD2_XENLA
9	15	1.8	872	1	CAD1_HUMAN
10	15	1.8	884	1	CAD1_MOUSE
11	15	1.8	886	1	CAD1_RAT
12	15	1.8	913	1	CAD4_CHICK
13	13	1.6	913	1	CAD4_MOUSE
14	13	1.6	916	1	CAD4_HUMAN
15	13	1.6	916	1	CAD4_MOUSE
16	12	1.4	877	1	CAD2_BOVIN
17	12	1.4	893	1	CAD2_BEARE
18	12	1.4	905	1	CADN_XENLA
19	12	1.4	906	1	CAD2_HUMAN
20	12	1.4	906	1	CAD2_MOUSE
21	12	1.4	906	1	CAD2_RAT
22	12	1.4	906	1	CAD2_XENLA
23	12	1.4	912	1	CAD2_CHICK
24	11	1.3	789	1	CAD6_RAT
25	11	1.3	789	1	CAD9_HUMAN
26	11	1.3	790	1	CCSA_CHLRE
27	10	1.2	353	1	CAD6_CHICK
28	10	1.2	790	1	CAD6_CHICK
29	9	1.1	381	1	CYB_DASBY
30	9	1.1	381	1	CYB_DASCR
31	9	1.1	413	1	ACD8_MOUSE
32	9	1.1	415	1	ACD8_HUMAN
33	9	1.1	788	1	CADA_HUMAN

34 9 1.1 790 1 CAD6_MOUSE
35 9 1.1 818 1 CDB1_HUMAN
36 9 1.1 3014 1 CLR1_HUMAN
37 9 1.1 3034 1 CLR1_MOUSE
38 8 1.0 93 1 GON2_RANCA
39 8 1.0 103 1 ENF1_CHICK
40 8 1.0 135 1 INL5_MOUSE
41 8 1.0 243 1 AETP_HAEIN
42 8 1.0 244 1 NXP2_MOUSE
43 8 1.0 262 1 NXP2_HUMAN
44 8 1.0 264 1 NXP2_BOVIN
45 8 1.0 265 1 CADA_MOUSE

ALIGNMENTS

RESULT 1
CAD3_HUMAN
ID CAD3_HUMAN STANDARD; PRT; 829 AA.
AC P22223;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cadherin-3 precursor (Placental-cadherin) (P-cadherin).
GN CDH3 OR CDHP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
GN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90003051; PubMed=2793940;
RA Shimoyama Y., Yoshida T., Terada M., Shimosato Y., Abe O.,
RA Hirohashi S.;
RT "Molecular cloning of a human Ca2+-dependent cell-cell adhesion
molecule homologous to mouse placental cadherin: its low expression
in human placental tissues.";
RL J. Cell Biol. 109:1787-1794(1989).
RN [2]
RP SEQUENCE FROM N.A.
RT TISSUE=Testis;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., S.W.,
FAHEY J., Heltan E., Kettman M., Madan A., Young A.C., Rodrigues S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson G.G.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield V.S.N., Kraywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP DISEASE.
RX MEDLINE=21470328; PubMed=11544476;
RA Sprecher E., Bergman R., Richard G., Lurie R., Shalev S., Cohen N.,
RA Petronius D., Shalata A., Ambinder Y., Leibin R., Periman I., Cohen N.,
RA Szargel R.;
RT "Hypotrichosis with juvenile macular dystrophy is caused by a
mutation in CDH3, encoding P-cadherin.";
RL Nat. Genet. 29:134-136(2001).
RN [4]

RP VARIANT HJMD HIS-503.
RX MEDLINE=22336734; PubMed=12445216;
RA Indelman M., Bergman R., Lurie R., Richard G., Miller B.,
RA Petronius D., Ciubutaro D., Leibur R., Sprecher E.;
RT "A missense mutation in CDH3, encoding P-cadherin, causes
RL hypotrichosis with juvenile macular dystrophy.";
RL J. Invest. Dermatol. 119:1210-1213(2002).
CC -1- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
CC They preferentially interact with themselves in a homophilic
CC manner in connecting cells; cadherins may thus contribute to the
CC sorting of heterogeneous cell types.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DISEASE: Defects in CDH3 are the cause of hypotrichosis with
CC juvenile macular dystrophy (HJMD) [MIM:601553]. HJMD is a rare
CC autosomal recessive disorder characterized by early hair loss
CC heralding severe degenerative changes of the retinal macula and
CC culminating in blindness during the second to third decade of
CC life.
CC -1- SIMILARITY: Contains 5 cadherin domains.
CC -1- DATABASE: NIMES=Mutations of the CDH3 gene;
CC NOTB=Retina International's Scientific Newsletter;
CC WWW="http://www.retina-international.com/sci-news/cdh3mut.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X63629; CAA45177.1; -;
CC EMBL: BC041846; AA41846.1; -;
CC F.R. A33659; IJHUCF.
CC HSP: P09803; ISUH.
CC Genew; HGNC:1762; CDH3.
CC MIM: 114021; -;
CC MIM: 601553; -;
CC GO: GO:0007155; P:cell adhesion; TAS.
CC InterPro: IPR002126; Cadherin.
CC InterPro: IPR002233; Cadherin_C_term.
CC Pfam: PF00028; cadherin; 5.
CC Pfam: PF01049; Cadherin_C_term; 1.
CC PRINTS: PR00205; CADHERIN.
CC SMART: SM00112; CA; 4.
CC PROSITE: PS00232; CADHERIN 1; 3.
CC PROSITE: PS00263; CADHERIN 2; 5.
CC Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
CC Signal; Disease mutation; Vision.
CC SIGNAL 1 24 POTENTIAL.
CC FT PROPEP 25 107
CC FT CHAIN 108 829
CC FT CADHERIN-3.
CC FT EXTRACELLULAR (POTENTIAL).
CC FT POTENTIAL 108 654
CC FT TRANSMEM 655 677
CC FT CYTOPLASMIC (POTENTIAL).
CC FT DOMAIN 678 829
CC FT CADHERIN 1.
CC FT DOMAIN 108 215
CC FT CADHERIN 2.
CC FT DOMAIN 216 328
CC FT CADHERIN 3.
CC FT DOMAIN 329 440
CC FT DOMAIN 441 546
CC FT CADHERIN 4.
CC FT DOMAIN 547 650
CC FT CADHERIN 5.
CC FT SER-RICH 785 800
CC FT DOMAIN 800 829
CC FT CARBOHYD 200 200
CC FT CARBOHYD 566 566
CC FT VARIANT 503 503
CC FT N-LINKED (GLNAC...) (POTENTIAL).
CC FT N-LINKED (GLNAC...) (POTENTIAL).
CC FT R -> H (in HJMD).
CC FT /FTID=VAR_015422.
CC FT 237 237
CC FT CONFLICT 237 263
CC FT H -> Q (IN REF. 2).
CC FT CONFLICT 563 563
CC FT SEQUENCE 829 AA; 91427 MW; E503CFEF5D981F1 CRC64;
CC
CC Query Match 100.0%; Score 829; DB 1; Length 829;
CC Best Local Similarity 100.0%; Pred. No. 0;
CC Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPRGPIASLLLOVLCWLOCAASEPCRAVFAEVTLEAGGAEPGQALGVKVGCPG 60
DB 1 MGLPRGPIASLLLOVLCWLOCAASEPCRAVFAEVTLEAGGAEPGQALGVKVGCPG 60
QY 61 QEPALFSTDMDDFTVRNGETVQERRSLKERNPLKIPPSKRILRRHKRDWVAPISVPENG 120
DB 61 QEPALFSTDMDDFTVRNGETVQERRSLKERNPLKIPPSKRILRRHKRDWVAPISVPENG 120
QY 121 KGPPFQRLNQLKSKNDRTKIFYSITGPGADSPGEGVFAVEKETGMLLNKPLDREBIAK 180
DB 121 KGPPFQRLNQLKSKNDRTKIFYSITGPGADSPGEGVFAVEKETGMLLNKPLDREBIAK 180
QY 181 YELFGHVAENGASVEDPMNISIVTDNDHKPKFTODTFRGSVLEGLVPGTSMVMTAT 240
DB 181 YELFGHVAENGASVEDPMNISIVTDNDHKPKFTODTFRGSVLEGLVPGTSMVMTAT 240
QY 241 DEDDAIYVYNGWAVSYHSQEPKDPHLMFTIHRSTGTISVISGLDREKVPYTLTQA 300
DB 241 DEDDAIYVYNGWAVSYHSQEPKDPHLMFTIHRSTGTISVISGLDREKVPYTLTQA 300
QY 301 TDMGDSGTTTAVAVVEILDANDNAPMFDQKYEAVPENAVGHEVORLTVTDLADNSP 360
DB 301 TDMGDSGTTTAVAVVEILDANDNAPMFDQKYEAVPENAVGHEVORLTVTDLADNSP 360
QY 361 AWRATYILMGDDGDHFTITTHPESNOGILITRKGLEDFAKNOHTLVVETNEAPFVKL 420
DB 361 AWRATYILMGDDGDHFTITTHPESNOGILITRKGLEDFAKNOHTLVVETNEAPFVKL 420
QY 421 PTSTATIWHVEDVNEAPVFPBPKVVEQGIPTGEPVCVYTAEDPKENQKISVILR 480
DB 421 PTSTATIWHVEDVNEAPVFPBPKVVEQGIPTGEPVCVYTAEDPKENQKISVILR 480
QY 481 DPAGWLAMPDPSGQVAVGLTDRDEQFVRNNIYEVVMVLAMDNGSPPTTGTGLTLTLD 540
DB 481 DPAGWLAMPDPSGQVAVGLTDRDEQFVRNNIYEVVMVLAMDNGSPPTTGTGLTLTLD 540
QY 541 VNDHGPVPEPRQITICNQSPVRHVLNITDKLSPHSFPQAOITDSDIYMTAEVNEEGD 600
DB 541 VNDHGPVPEPRQITICNQSPVRHVLNITDKLSPHSFPQAOITDSDIYMTAEVNEEGD 600
QY 601 TVVLSLKKFLKQDIDYVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGFWKGFILPVL 660
DB 601 TVVLSLKKFLKQDIDYVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGFWKGFILPVL 660
QY 661 GAVLALLFLVLLVLLVLRKKRIKEPLLPEDDTRDNVYFYGEGGGEEDQDIDITQHR 720
DB 661 GAVLALLFLVLLVLLVLRKKRIKEPLLPEDDTRDNVYFYGEGGGEEDQDIDITQHR 720
QY 721 GLEARPEVVLNDVAPITPTPMYRPRPANDPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
DB 721 GLEARPEVVLNDVAPITPTPMYRPRPANDPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
QY 781 DYEGSGSDAASLSLTSASDQDQDYDYNWGSRFKFLADMYGGGDD 829
DB 781 DYEGSGSDAASLSLTSASDQDQDYDYNWGSRFKFLADMYGGGDD 829

RESULT 2
CAD3_BOVIN
ID CAD3_BOVIN STANDARD; PRT; 491 AA.
AC P19535;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cadherin-3 (Placental-cadherin) (P-cadherin) (Fragment).
GN CDH3 OR CDHP.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.

```

RX MEDLINE=90360979; PubMed=2390969;
RA Liaw C.W., Cannon C., Power M.D., Kiboneka P.K., Rubin L.L.;
RT "Identification and cloning of two species of cadherins in bovine
RL endothelial cells.";
CC EMBL J. 9:2701-2708(1990).
CC -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
CC They preferentially interact with themselves in a homophilic
CC manner in connecting cells; cadherins may thus contribute to the
CC sorting of heterogeneous cell types.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- SIMILARITY: Contains 5 cadherin domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X53614; CAA37676.1; -.
CC PIR: S11694; IJBOCP.
CC InterPro: IPR002126; Cadherin.
CC InterPro: IPR000233; Cadherin_C_term.
CC Pfam: PF00028; cadherin; 3.
CC Pfam: PF01049; Cadherin_C term; 1.
CC PRINTS: PR00205; CADHERIN.
CC SMART: SM00112; CA; 2
CC PROSITE: PS00232; CADHERIN 1; 1.
CC PROSITE: PS00268; CADHERIN 2; 2.
CC Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
CC KW NON_TER 1 1
CC FT DOMAIN <1 316 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 317 339 POTENTIAL.
CC FT CYTOPLASM 340 491 CYTOPLASMIC (POTENTIAL).
CC FT DOMAIN <1 102 CADHERIN 3.
CC FT DOMAIN 103 208 CADHERIN 4.
CC FT DOMAIN 209 314 CADHERIN 5.
CC FT DOMAIN 447 462 SER-RICH.
CC FT CARBOHYD 228 228 N-LINKED (GLCNAC...) (POTENTIAL).
CC SQ SEQUENCE 491 AA; 54207 MW; 6BE80A5918C4771 CRC64;

Query Match 7.4%; Score 61; DB 1; Length 491;
Best Local Similarity 100.0%; Pred. No. 7.4e-49;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 676 LVKFKKIKPELLPDDTRDNVFFYGGGGEDQDYDITQHRGLRPEVVLNDVA 735
Db 338 LVKFKKIKPELLPDDTRDNVFFYGGGGEDQDYDITQHRGLRPEVVLNDVA 397

QY 736 P 736
Db 398 P 398

RESULT 3
CAD3_MOUSE STANDARD; PRT; 822 AA.
AC P10287; Q61465;
AT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cadherin-3 precursor (Placental-cadherin) (P-cadherin).
GN CDH3 OR CDHP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8811554; PubMed=3428270;
RA Nose A., Nagatuchi A., Takeichi M.;
RT "Isolation of placental cadherin cDNA: Identification of a novel gene

```

```

RT family of cell-cell adhesion molecules.";
RL EMBL J. 6:3655-3661(1987).
RN [2]
RP PARTIAL SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Embryo;
RX MEDLINE=91360343; PubMed=1886768;
RA Hatta M., Miyatani S., Copeland N.G., Gilbert D.J., Jenkins N.A.,
RA Takeichi M.;
RT "Genomic organization and chromosomal mapping of the mouse P-cadherin
RT gene.";
RL Nucleic Acids Res. 19:4437-4441(1991).
RN [3]
RP SEQUENCE OF 1-55 FROM N.A.
RC STRAIN=C57BL/6;
RA Hatta M., Takeichi M.;
RT "Complex cell type-specific transcriptional regulation by the promoter
RT and an intron of the mouse P-cadherin gene.";
RL Dev. Growth Differ. 36:509-519(1994).
RN [4]
RP DEVELOPMENTAL STAGE.
RC STRAIN=C57BL/6; TISSUE=Testis;
RX MEDLINE=97033837; PubMed=8879495;
RA Munro S.B., Blaschuk O.W.;
RT "A comprehensive survey of the cadherins expressed in the testes of
RT fetal, immature, and adult mice utilizing the polymerase chain
RT reaction.";
RL Biol. Reprod. 55:822-827(1996).
CC -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
CC They preferentially interact with themselves in a homophilic
CC manner in connecting cells; cadherins may thus contribute to the
CC sorting of heterogeneous cell types.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- DEVELOPMENTAL STAGE: Expression is high in both fetal and newborn
CC testis but minimal in testis of 7-day-old animals. Not detected in
CC testis of 21-day-old or adult.
CC -!- SIMILARITY: Contains 5 cadherin domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X06340; CAA29646.1; -.
CC EMBL: D12688; EBA02186.1; -.
CC PIR: S03183; IJMSCP.
CC HSP: P09803; 1SUH.
CC MGD; MGI:88356; Cdh3.
CC GO; GO:0005886; C:plasma membrane; IDA.
CC InterPro: IPR002126; Cadherin.
CC InterPro: IPR000233; Cadherin_C_term.
CC Pfam: PF00028; cadherin; 5.
CC Pfam: PF01049; Cadherin_C term; 1.
CC PRINTS: PR00205; CADHERIN.
CC SMART: SM00112; CA; 4.
CC PROSITE: PS00232; CADHERIN 1; 3.
CC PROSITE: PS00268; CADHERIN 2; 4.
CC Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
CC Signal.
CC SIGNAL 1 25 POTENTIAL.
CC PROPEP 26 99
CC CHAIN 100 822 CADHERIN-3.
CC DOMAIN 100 647 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 648 670 POTENTIAL.
CC DOMAIN 671 822 CYTOPLASMIC (POTENTIAL).
CC DOMAIN 100 207 CADHERIN 1.
CC DOMAIN 208 320 CADHERIN 2.
CC DOMAIN 321 432 CADHERIN 3.
CC DOMAIN 433 538 CADHERIN 4.
CC DOMAIN 539 645 CADHERIN 5.
CC DOMAIN 778 793 SER-RICH.

```

FT CARBOHYD 192 192 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 558 558 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 822 AA; 90754 MW; 8497502115B14DC3 CRC64;

Query Match 6.2%; Score 51; DB 1; Length 822;
 Best Local Similarity 100.0%; Pred. No. 2.5e-39;
 Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 684 KEPLLPEDTRDNVYFYGEGGEGEDQYDITQLHGLEARPEVVLNDV 734
 DB 677 KEPLLPEDTRDNVYFYGEGGEGEDQYDITQLHGLEARPEVVLNDV 727

RESULT 4
 ID -CAD3_PIG STANDARD; PRT; 145 AA.

AC O18926; 2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Cadherin-3 (Placental-cadherin) (P-cadherin) (Fragment).
 GN CDH3
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.

RC TISSUE=Retinal pigment epithelium;
 RA Lutz D.A., Zheng J.J.;
 RT "Expression of multiple cadherins in adult retinal pigment epithelial (RPE) cells";
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins. They preferentially interact with themselves in a homophilic manner in connecting cells; cadherins may thus contribute to the sorting of heterogeneous cell types.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
 CC -!- SIMILARITY: Contains 5 cadherin domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AF033826; AAB87087.1; -;
 DR HSSP; P09803; 1SUH.
 DR InterPro; IPR002126; Cadherin.
 DR Pfam; PF00028; cadherin; 1.
 DR PRINTS; PS00205; CADHERIN.
 DR SMART; SM00112; CA; 1.
 DR PROSITE; PS00232; CADHERIN 1; 1.
 DR PROSITE; PS0268; CADHERIN 2; 1.
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
 FT NON_TER 1 1
 FT DOMAIN <1>145 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN <1>39 CADHERIN 1.
 FT DOMAIN 40 >145 CADHERIN 2.
 FT CARBOHYD 24 24
 FT NON_TER 145 145
 SQ SEQUENCE 145 AA; 15617 MW; 8AA2C49E76EB40EC CRC64;

Query Match 3.3%; Score 27; DB 1; Length 145;
 Best Local Similarity 100.0%; Pred. No. 1.8e-17;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 FRGSVLEGLVPGTSVMQVATDEDDAI 246
 DB 44 FRGSVLEGLVPGTSVMQVATDEDDAI 70

RESULT 5
 ID -CADL_CHICK STANDARD; PRT; 732 AA.
 AC P33145;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE B-cadherin precursor (K-CAM protein) (Fragment).
 GN K-CAM.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92107987; PubMed=1763068;
 RA Sorkin B.C., Gallin W.J., Edelman G.M., Cunningham B.A.;
 RT "Genes for two calcium-dependent cell adhesion molecules have similar structures and are arranged in tandem in the chicken genome.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:11545-11549 (1991).
 RN [2]
 RP SEQUENCE OF 7-732 FROM N.A.
 RC TISSUE=Embryonic brain;
 RX MEDLINE=91225083; PubMed=2026653;
 RA Napolitano E.W., Venstrom K., Wheeler E.F., Reichardt L.F.;
 RT "Molecular cloning and characterization of B-cadherin, a novel chick cadherin";
 RL J. Cell Biol. 113:893-905 (1991).
 CC -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins. They preferentially interact with themselves in a homophilic manner in connecting cells; cadherins may thus contribute to the sorting of heterogeneous cell types. B-cadherin may have important functions in neurogenesis, in at least some epithelia, and in embryogenesis.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- TISSUE SPECIFICITY: Expressed in a wide variety of tissues.
 CC -!- SIMILARITY: Contains 5 cadherin domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; M81894; AAA48929.1; -;
 DR EMBL; X58518; CAA41408.1; -;
 DR PIR; A41634; IJCHCB.
 DR HSSP; P09803; 1SUH.
 DR InterPro; IPR002126; Cadherin.
 DR InterPro; IPR00233; Cadherin_C-term.
 DR Pfam; PF00028; cadherin; 5.
 DR Pfam; PF01049; Cadherin_C-term; 1.
 DR PRINTS; PR00205; CADHERIN.
 DR PROSITE; PS00232; CADHERIN 1; 3.
 DR PROSITE; PS0268; CADHERIN 2; 5.
 DR Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
 FT NON_TER 1 1
 FT PROPEP <1>6
 FT CHAIN 7 732
 FT DOMAIN 6 554
 FT TRANSMEM 555 580
 FT DOMAIN 581 732
 FT DOMAIN 6 114
 FT DOMAIN 115 227
 FT DOMAIN 228 339
 FT DOMAIN 340 443
 FT DOMAIN 444 554
 FT DOMAIN 554 702
 FT CARBOHYD 137 137

FT POTENTIAL.
 FT B-CADHERIN.
 FT EXTRACELLULAR (POTENTIAL).
 FT POTENTIAL.
 FT CYTOPLASMIC (POTENTIAL).
 FT CADHERIN 1.
 FT CADHERIN 2.
 FT CADHERIN 3.
 FT CADHERIN 4.
 FT CADHERIN 5.
 FT SER-RICH.
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).

```

FT CARBOHYD 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 414 414 M -> V (IN REF. 2).
SQ SEQUENCE 732 AA; 80613 MW; 091D59A6A16CBD45 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 1.5e-07; Length 732;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 797 SSASDQDDYDLNHWG 813
Db 699 SSASDQDDYDLNHWG 715

RESULT 6
CAD1_CHICK
ID _CAD1_CHICK STANDARD; PRT; 887 AA.
AC P08641;
DT 01-AUG-1998 (Rel. 08, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Epithelial-cadherin precursor (E-cadherin) (Cadherin-1) (Liver cell
DE adhesion molecule) (L-CAM).
GN CDH1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89017248; PubMed=3174655;
RA Sorkin B.C., Hemperly J.J., Edelman G.M., Cunningham B.A.;
RT "Structure of the gene for the liver cell adhesion molecule, L-CAM.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:7617-7621(1988).
RN [2]
RP SEQUENCE OF 51-887 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=87204217; PubMed=3472238;
RA Gallin W.J., Sorkin B.C., Edelman G.M., Cunningham B.A.;
RT "Sequence analysis of a cDNA clone encoding the liver cell adhesion
RT molecule, L-CAM.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:2808-2812(1987).
CC -1- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
CC They preferentially interact with themselves in a homophilic
CC manner in connecting cells; cadherins may thus contribute to the
CC sorting of heterogeneous cell types. E-cadherin is a ligand for
CC integrin alpha-E/beta-7.
CC -1- SUBUNIT: Homodimer.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: Non-neural epithelial tissues.
CC -1- SIMILARITY: Contains 5 cadherin domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M22190; AAA82572.1; JOINED.
DR EMBL; M22180; AAA82572.1; JOINED.
DR EMBL; M22181; AAA82572.1; JOINED.
DR EMBL; M22182; AAA82572.1; JOINED.
DR EMBL; M22186; AAA82572.1; JOINED.
DR EMBL; M22183; AAA82572.1; JOINED.
DR EMBL; M22194; AAA82572.1; JOINED.
DR EMBL; M22184; AAA82572.1; JOINED.
DR EMBL; M22185; AAA82572.1; JOINED.
DR EMBL; M22189; AAA82572.1; JOINED.
DR EMBL; M22193; AAA82572.1; JOINED.
DR EMBL; M22187; AAA82572.1; JOINED.
DR EMBL; M22192; AAA82572.1; JOINED.
DR EMBL; M22191; AAA82572.1; JOINED.

```

```

DR EMBL; M22195; AAA82572.1; JOINED.
DR EMBL; M16260; AAA82573.1; .
DR PIR; A30201; IJCHCL.
DR HSSP; P09803; 1SUH.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR000333; Cadherin_C_term.
DR Pfam; PF00028; cadherin; 5.
DR Pfam; PF01049; Cadherin_C_term; 1.
DR PRINTS; P00205; CADHERIN.
DR SMART; SM00112; CA; 4.
DR PROSITE; PS00232; CADHERIN_1; 3.
DR PROSITE; PS0268; CADHERIN_2; 5.
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
KW Signal.
FT SIGNAL 1 26 POTENTIAL.
FT PROPEP 27 160
FT CHAIN 161 887
FT DOMAIN 161 714
FT TRANSMEM 715 735
FT DOMAIN 736 887
FT DOMAIN 161 288
FT DOMAIN 269 381
FT DOMAIN 382 493
FT DOMAIN 494 599
FT DOMAIN 600 704
FT DOMAIN 844 855
FT CARBOHYD 291 291
FT CARBOHYD 346 346
FT CARBOHYD 564 564
FT CARBOHYD 643 643
FT CONFLICT 140 140
SQ SEQUENCE 887 AA; 97783 MW; F56A8A9779A94A40 CRC64;

```

Query Match 2.1%; Score 17; DB 1; Length 887;
 Best Local Similarity 100.0%; Pred. No. 1.8e-07;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 797 SSASDQDDYDLNHWG 813
Db 854 SSASDQDDYDLNHWG 870

```

```

RESULT 7
CADF_XENLA
ID _CADF_XENLA STANDARD; PRT; 880 AA.
AC P31148; O91543;
DT 01-OCT-1993 (Rel. 27, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE EP-cadherin precursor (C-cadherin).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xeropodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91372132; PubMed=1893866;
RA Ginsberg D., Deslmon D., Geiger B.;
RT "Expression of a novel cadherin (EP-cadherin) in unfertilized eggs
RT and early Xenopus embryos.";
RN [2]
RP SEQUENCE FROM N.A.
RA Lee C.H.;
RN [3]
RP REVISION TO 260.
RA Flament S.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
CC They preferentially interact with themselves in a homophilic
CC manner in connecting cells; cadherins may thus contribute to the

```


QY 699 FTYGEGGGEDDYYD 714
Db 752 FTYGEGGGEDDYYD 767

RESULT 9
CAD1_XENLA STANDARD; PRT; 872 AA.
AC P30944; Q91709;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Epithelial-cadherin precursor (E-cadherin) (Uvomorulin) (XTCAD-1).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
CX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95324376; PubMed=7600966;
RA Levine E., Lee C.H., Kintner C., Gumbiner B.M.;
RT "Selective disruption of E-cadherin function in early Xenopus embryos
by a dominant negative mutant.";
RL Development 120:901-909(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC Tissue=tail bud;
RX MEDLINE=94368839; PubMed=8086449;
RA Tooi O., Fujii G., Tashiro K., Shikawa K.;
RT "Molecular cloning of cDNA for XTCAD-1, a novel Xenopus cadherin, and
its expression in adult tissues and embryos of Xenopus laevis.";
RL Biochim. Biophys. Acta 1219:121-128(1994).
RN [3]
RP SEQUENCE OF 149-872 FROM N.A.
RX MEDLINE=94363396; PubMed=8081882;
RA Broders F., Girault J.M., Simonneau L., Thiery J.P.;
RT "Sequence and distribution of Xenopus laevis E-cadherin transcripts.";
RL Cell Adhes. Commun. 1:265-277(1993).
RN [4]
RP SEQUENCE OF 149-169.
RX MEDLINE=91347911; PubMed=1879345;
RA Angres E., Mueller A.H.J., Kellermann J., Hausen P.;
RT "Differential expression of two cadherins in Xenopus laevis.";
RL Development 111:829-844(1991).
CC -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
CC They preferentially interact with themselves in a homophilic
CC manner in connecting cells; cadherins may thus contribute to the
CC sorting of heterogeneous cell types. E-cadherin is a ligand for
CC integrin alpha-E/beta-7.
CC -!- SUBUNIT: Homodimer.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: Non-neural epithelial tissues.
CC -!- DEVELOPMENTAL STAGE: Appears in the embryonic ectoderm during
CC gastrulation when epidermal differentiation commences and it
CC disappears from the neural plate area upon neural induction.
CC -!- SIMILARITY: Contains 5 cadherin domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U04708; AAA93116.1; -
CC EMBL; L29057; AAA61489.1; -
CC EMBL; X75454; CAA53206.1; -
CC HSSP; P09803; ISUH.
CC InterPro; IPR002126; Cadherin.
CC InterPro; IPR000233; Cadherin_C_term.
PFam; PF00028; cadherin; 5.

DR PFam; PF01049; Cadherin_C_term; 1.
DR PRINTS; PRO0205; CADHERIN.
DR SMART; SM00112; CA; 4.
DR PROSITE; PS00232; CADHERIN_1; 3.
DR PROSITE; PS00268; CADHERIN_2; 4.
KW Cell adhesion; Glycoprotein; transmembrane; Calcium-binding; Repeat;
KW Signal.
FT SIGNAL. 1 25 POTENTIAL.
FT PROPEP 26 148
FT CHAIN 149 872 EPITHELIAL-CADHERIN.
FT DOMAIN 149 701 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 702 722 POTENTIAL.
FT DOMAIN 723 872 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 244 254 CADHERIN 1.
FT DOMAIN 358 368 CADHERIN 2.
FT DOMAIN 577 587 CADHERIN 3.
FT DOMAIN 710 721 POLY-LEU.
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 456 456 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 552 552 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 669 669 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 242 242 I -> V (IN REF. 3).
FT CONFLICT 332 332 N -> T (IN REF. 3).
FT CONFLICT 487 487 E -> R (IN REF. 3).
FT CONFLICT 501 502 AT -> CS (IN REF. 3).
FT CONFLICT 539 541 GNG -> EMA (IN REF. 3).
FT CONFLICT 548 548 K -> R (IN REF. 3).
FT CONFLICT 557 557 V -> G (IN REF. 3).
FT CONFLICT 567 567 P -> L (IN REF. 3).
FT CONFLICT 602 604 GFR -> EPQ (IN REF. 3).
FT CONFLICT 638 640 GQS -> DK (IN REF. 2).
FT CONFLICT 640 647 SILERPK -> VYLSSDL (IN REF. 3).
FT CONFLICT 660 660 T -> A (IN REF. 3).
FT CONFLICT 770 770 A -> S (IN REF. 3).
FT CONFLICT 842 842 P -> S (IN REF. 2 AND 3).
FT CONFLICT 870 871 DE -> GED (IN REF. 3).
SQ SEQUENCE 872 AA; 96065 MW; 08637967EFPB4664 CRC64;

Query Match 1.8%; Score 15; DB 1; Length 872;
Best Local Similarity 100.0%; Pred. No. 1.3e-05;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 744 YRPRPNPDEIGNFI 758

Db 788 YRPRPNPDEIGNFI 802

RESULT 10

CAD1_HUMAN
ID CAD1_HUMAN STANDARD; PRT; 882 AA.
AC P12830; Q13799; Q14216; Q15855; Q16194;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-JUL-1993 (Rel. 28, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Epithelial-cadherin precursor (E-cadherin) (Uvomorulin) (Cadherin-1)
DE (CAM 120/80).
GN CDH1 OR UVO OR CDHE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=92211394; PubMed=8459805;
RA Bussemakers M.J.G., Mees S.G.M., van Bokhoven A., Debruyne F.M.J.,
RA Schaiken J.A.;
RT "Molecular cloning and characterization of the human E-cadherin
cDNA.";
RL Mol. Biol. Rep. 17:123-128(1993).
RN [2]
RP SEQUENCE FROM N.A.

RA Kelker W., Warda A., Oda T., Hirohashi S., Kemler R., Birchmeier W.;
 RA Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RN SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Rimm D.L., Morrow J.S.;
 RA "Molecular cloning of human E-cadherin suggests a novel subdivision of
 the cadherin superfamily";
 RT Biochem. Biophys. Res. Commun. 200:1754-1761(1994).
 RN [4]
 RN SEQUENCE OF 56-882 FROM N.A.
 RC TISSUE=Placenta;
 RA MEDLINE=95324290; PubMed=7601454;
 RA Berx G., Staes K., van Hengel J., Molemans F., Bussemakers M.J.G.,
 van Bokhoven A., van Roy F.;
 RA "Cloning and characterization of the human invasion suppressor gene
 E-cadherin (CDH1)";
 RT Genomics 26:281-289(1995).
 RN [5]
 RN SEQUENCE OF 172-311 FROM N.A.
 RC TISSUE=Liver;
 RA Mansouri A., Spurr N., Goodfellow P.N., Kemler R.;
 RA "Characterization and chromosomal localization of the gene encoding
 the human cell adhesion molecule uvomorulin";
 RT Differentiation 38:67-71(1988).
 RN [6]
 RN SEQUENCE OF 265-392 FROM N.A.
 RC TISSUE=Liver;
 RA Priden U.H.;
 RA Submitted (MAR-1990) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RN SEQUENCE OF 1-16 FROM N.A.
 RC MEDLINE=94380041; PubMed=8093045;
 RA Bussemakers M.J.G., Girolodi L.A., van Bokhoven A., Schalken J.A.;
 RA "Transcriptional regulation of the human E-cadherin gene in human
 prostate cancer cell lines: characterization of the human E-cadherin
 gene promoter";
 RT Biochem. Biophys. Res. Commun. 203:1284-1290(1994).
 RN [8]
 RN SEQUENCE OF 1-16 FROM N.A.
 RC TISSUE=Placenta;
 RA Yashima K., Kanai Y., Ochiai A., Shimoyama Y., Sugimura T.,
 Hirohashi S.;
 RA "Silencing of the E-cadherin invasion-suppressor gene by CpG
 methylation in human carcinomas";
 RT Proc. Natl. Acad. Sci. U.S.A. 92:7416-7419(1995).
 RN [9]
 RN INTERCHAIN DISULFIDE BOND.
 RA MEDLINE=21975237; PubMed=11856785;
 RA Makgansar I.T., Nguyen P.D., Ikseue A., Kuczerka K., Dentler W.,
 Urbauer J.L., Galeva N., Alterman M., Sahaan T.J.;
 RA "Disulfide bond formation promotes the cis- and trans-dimerization of
 the E-cadherin-derived first repeat";
 RT J. Biol. Chem. 277:16002-16010(2002).
 RN [10]
 RN DISEASE
 RA MEDLINE=99406500; PubMed=10477433;
 RA Guilford P.J., Hopkins J.B.W., Grady W.M., Markowitz S.D., Willis J.,
 Lynch H., Rajput A., Wiesner G.L., Lindor N.M., Burgart L.J.,
 Toro T.T., Lee D., Limacher J.-M., Shaw D.W., Findlay M.P.N.,
 Reeve A.E.;
 RA "E-cadherin germline mutations define an inherited cancer syndrome
 dominated by diffuse gastric cancer";
 RT Hum. Mutat. 14:249-255(1999).
 RN [11]
 RN REVIEW ON VARIANTS.
 RA MEDLINE=98415721; PubMed=9744472;
 RA Berx G., Becker K.-F., Hoefler H., van Roy F.;
 RA "Mutations of the human E-cadherin (CDH1) gene";
 RT Hum. Mutat. 12:226-237(1998).
 RN [12]
 RN SEQUENCE OF 337-476 FROM N.A., AND VARIANTS ALA-370 AND ASP-473.
 RA MEDLINE=94306394; PubMed=8033105;
 RA Becker K.-F., Atkinson M.J., Reich U., Becker I., Nekarda H.,
 Siewert J.R., Hoefler H.;
 RA "E-cadherin gene mutations provide clues to diffuse type gastric
 carcinomas";
 RT Cancer Res. 54:3845-3852(1994).
 RN [13]
 RN VARIANT LOBULAR BREAST CARCINOMA SER-315.
 RA MEDLINE=95049851; PubMed=7961105;
 RA Kanai Y., Oda T., Tsuda H., Ochiai A., Hirohashi S.;
 RA "Point mutation of the E-cadherin gene in invasive lobular carcinoma
 of the breast";
 RT Jpn. J. Cancer Res. 85:1035-1039(1994).
 RN [14]
 RN VARIANTS GYNECOLOGIC CANCERS THR-617; VAL-711 AND GLY-838.
 RA MEDLINE=94355985; PubMed=8075649;
 RA Risinger J.I., Berchuck A., Kohler M.F., Boyd J.;
 RA "Mutations of the E-cadherin gene in human gynecologic cancers";
 RT Nat. Genet. 7:98-102(1994).
 RN [15]
 RN VARIANT GASTRIC ADENOCARCINOMA 274-GLY--PRO-277 DEL.
 RA MEDLINE=94173928; PubMed=8127895;
 RA Oda T., Kanai Y., Oyama T., Yoshiura K., Shimoyama Y., Birchmeier W.,
 Sugimura T., Hirohashi S.;
 RA "E-cadherin gene mutations in human gastric carcinoma cell lines";
 RT Proc. Natl. Acad. Sci. U.S.A. 91:1858-1862(1994).
 RN [16]
 RN VARIANT GASTRIC CARCINOMA PRO-193.
 RA MEDLINE=96390918; PubMed=8797891;
 RA Muta H., Noguchi M., Kanai Y., Ochiai A., Nawata H., Hirohashi S.;
 RA "E-cadherin gene mutations in signet ring cell carcinoma of the
 stomach";
 RT Jpn. J. Cancer Res. 87:843-848(1996).
 RN [17]
 RN VARIANTS GASTRIC CARCINOMA ASP-400 DEL AND 418-ASP--PHE-423 DEL.
 RA MEDLINE=97197648; PubMed=9045944;
 RA Tamura G., Sakata K., Nishizuka S., Maesawa C., Suzuki Y., Iwaya T.,
 Terashima M., Saito K., Satodate R.;
 RA "Inactivation of the E-cadherin gene in primary gastric carcinomas and
 gastric carcinoma cell lines";
 RT Jpn. J. Cancer Res. 87:1153-1159(1996).
 RN [18]
 RN VARIANT THYROID TUMOR THR-592.
 RA MEDLINE=97138061; PubMed=8985087;
 RA Soares P., Berx G., van Roy F., Sobrinho-Simoes M.;
 RA "E-cadherin gene alterations are rate events in thyroid tumors";
 RT Int. J. Cancer 70:32-38(1997).
 RN [19]
 RN VARIANTS ASP-336 AND ILE-470.
 RA MEDLINE=98196671; PubMed=9537325;
 RA Guilford P.J., Hopkins J.B.W., Harraway J., McLeod M., McLeod N.,
 Harawira P., Taite H., Scoular R., Miller A., Reeve A.E.;
 RA "E-cadherin germline mutations in familial gastric cancer";
 RT Nature 392:402-405(1998).
 RN [20]
 RN VARIANTS HDGC GLY-244 AND ALA-487.
 RA MEDLINE=99253140; PubMed=10319582;
 RA Yoon K.-A., Xu J.-L., Yang H.-K., Kim W.H., Park S.Y., Park J.-G.;
 RA "Germline mutations of E-cadherin gene in Korean familial gastric
 cancer patients";
 RT J. Hum. Genet. 44:177-180(1999).
 RN [21]
 RN VARIANT COLORECTAL CANCER ALA-340.
 RA MEDLINE=20357134; PubMed=10896919;
 RA Kim H.C., Wheeler J.M.D., Kim J.C., Ilyas M., Beck N.E., Kim B.S.,
 Park K.C., Bodmer W.F.;
 RA "The E-cadherin gene (CDH1) variants T340A and L599V in gastric and
 colorectal cancer patients in Korea";
 RT Gut 47:262-267(2000).
 RN [22]
 RN VARIANT ALA-270.


```

RX MEDLINE=21562463; PubMed=11705864;
RA Ikonen T., Matikainen M., Mononen N., Hyytinen E.R., Helin H.J.,
RA Tomola S., Rammela T.L., Pukkala E., Schleutker J., Kallioniemi O.P.,
RA Koivisto P.A.;
RT "Association of E-cadherin germ-line alterations with prostate
RT cancer.";
RL Clin. Cancer Res. 7:3465-3471(2001).
RN [23]
RP VARIANT THR-592.
RX MEDLINE=21446527; PubMed=11562785;
RA Salahshor S., Hou H., Diep C.B., Loukola A., Zhang H., Liu T.,
RA Chen J., Iselius L., Rubio C., Lothe R.A., Aaltonen L., Sun X.F.,
RA Lindmark G., Lindblom A.;
RT "A germline E-cadherin mutation in a family with gastric and colon
RT cancer.";
RL Int. J. Mol. Med. 8:439-443(2001).
RN [24]
RP VARIANT HDGC ALA-340.
RX MEDLINE=22021361; PubMed=11968083;
RA Oliveira C., Bordin M.C., Grehn N., Huntsman D., Suriano G.,
RA Machado J.C., Kiviluoto T., Aaltonen L., Jackson C.E., Seruca R.,
RA Caldas C.;
RT "Screening E-cadherin in gastric cancer families reveals germline
RT mutations only in hereditary diffuse gastric cancer kindred.";
RL Hum. Mutat. 19:510-517(2002).
RN [6]
CC -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
CC They preferentially interact with themselves in a homophilic
CC manner in connecting cells; cadherins may thus contribute to the
CC sorting of heterogeneous cell types. E-cadherin has a potent
CC invasive suppressor role. It is also a ligand for integrin alpha-
CC E/beta-7.
CC

Query Match 1.0%; Score 15; DB 1; Length 882;
Best Local Similarity 100.0%; Pred. No. 1.3e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 815 RPKKLADMYGGGDD 829
DB 868 RPKKLADMYGGGDD 882
|||||
RESULT 11
CAD1_MOUSE STANDARD; PRT; 884 AA.
AC P09803; Q61377;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Epithelial-cadherin precursor (E-cadherin) (Uvomorulin) (Cadherin-1)
DE (ARC-1).
GN CDH1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR;
RX MEDLINE=87315445; PubMed=3498123;
RA Nagatuchi A., Shikayoshi Y., Okazaki K., Yasuda K., Takeichi M.;
RT "Transformation of cell adhesion properties by exogenously introduced
RT E-cadherin cDNA.";
RL Nature 329:341-343(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/Sv;
RX MEDLINE=92093614; PubMed=1754391;
RA Ringwald M., Baribault H., Schmidt C., Kemler R.;
RT "The structure of the gene coding for the mouse cell adhesion
RT molecule uvomorulin".
RL Nucleic Acids Res. 19:6533-6539(1991).
RN [3]
RP SEQUENCE OF 174-884 FROM N.A., AND SEQUENCE OF 157-181.

```

```

RX MEDLINE=88111553; PubMed=3501370;
RA Ringwald M., Schuh R., Vestweber D., Eistetter H., Lottspeich F.,
RA Engel J., Doelz R., Jaehnig F., Epplen J., Mayer S., Mueller C.,
RA Kemler R.;
RT "The structure of cell adhesion molecule uvomorulin. Insights into
RT the molecular mechanism of Ca2+-dependent cell adhesion.";
RL EMBO J. 6:3647-3653(1987).
RN [4]
RP SEQUENCE OF 1-15 FROM N.A.
RX MEDLINE=92107977; PubMed=1763063;
RA Behrens J., Loewrick O., Klein-Hitpass L., Birchmeier W.;
RT "The E-cadherin promoter: functional analysis of a G.C-rich region
RT and an epithelial cell-specific palindromic regulatory element.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:11495-11499(1991).
RN [5]
RP DEVELOPMENTAL STAGE.
RC STRAIN=C57BL/6; TISSUE=Testis;
RX MEDLINE=97033837; PubMed=8879495;
RA Munro S.B., Blaschuk O.W.;
RT "A comprehensive survey of the cadherins expressed in the testes of
RT fetal, immature, and adult mice utilizing the polymerase chain
RT reaction.";
RL Biol. Reprod. 55:822-827(1996).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 157-370.
RX MEDLINE=96176249; PubMed=8598933;
RA Nagar B., Overduin W., Ikura M., Rini J.M.;
RT "Structural basis of calcium-induced E-cadherin rigidification and
RT dimerization.";
RL Nature 380:360-364(1996).
RN [7]
RP STRUCTURE BY NMR OF 157-260.
RX MEDLINE=96271285; PubMed=8785495;
RA Overduin M., Tong K.I., Kay C.M., Ikura M.;
RT "1H, 15N and 13C resonance assignments and monomeric structure of the
RT amino-terminal extracellular domain of epithelial cadherin.";
RL J. Biomol. NMR 7:173-189(1996).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS) OF 577-728 IN COMPLEX WITH
RP CTNNB1 AND PHOSPHORYLATION.
RX MEDLINE=21246507; PubMed=11348595;
RA Huber A.H., Weis W.I.;
RT "The structure of the beta-catenin/E-cadherin complex and the
RT molecular basis of diverse ligand recognition by beta-catenin.";
RL Cell 105:391-402(2001).
CC -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
CC They preferentially interact with themselves in a homophilic
CC manner in connecting cells; cadherins may thus contribute to the
CC sorting of heterogeneous cell types. E-cadherin is a ligand for
CC integrin alpha-E/beta-7.
CC -!- SUBUNIT: Homodimer. Binds CTNNB1.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: Non-neural epithelial tissues.
CC -!- DEVELOPMENTAL STAGE: In the testis, expression is highest in fetal
CC gonad, then decreases 5-fold in newborn. Detectable in 7-day-old
CC but not in 21-day-old or adult.
CC -!- SIMILARITY: Contains 5 cadherin domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sb-sib.ch).
CC
CC EMBL; X06115; CAA29488.1; -
CC EMBL; X60961; CAA43292.1; -
CC EMBL; X60962; CAA43292.1; JOINED.
CC EMBL; X60963; CAA43292.1; JOINED.
CC EMBL; X60964; CAA43292.1; JOINED.
CC EMBL; X60965; CAA43292.1; JOINED.
CC EMBL; X60966; CAA43292.1; JOINED.

```



```
CC EMBL; AB017696; BAA84920.1; -
CC EMBL; AF177680; AAR87055.1; -
CC EMBL; AJ000540; CAA04173.1; -
CC HSPF; P03903; IEDH.
CC InterPro; IPR002126; Cadherin.
CC InterPro; IPR002333; Cadherin_C_term.
CC Pfam; PF00028; cadherin; 5.
CC Pfam; PF01049; Cadherin_C_term; 1.
CC PRINTS; PR00205; CADHERIN.
CC SMART; SM00112; CA; 4.
CC PROSITE; PS00232; CADHERIN_1; 3.
CC PROSITE; PS0268; CADHERIN_2; 5.
CC Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
KW Signal; Phosphorylation.
FT SIGNAL 1 23 POTENTIAL.
FT PROPEP 24 158 POTENTIAL.
FT CHAIN 159 886 EPIHELIAL-CADHERIN.
FT DOMAIN 24 713 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 714 734 POTENTIAL.
FT DOMAIN 735 886 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 1 266 CADHERIN_1.
FT DOMAIN 267 379 CADHERIN_2.
FT DOMAIN 380 490 CADHERIN_3.
FT DOMAIN 491 597 CADHERIN_4.
FT DOMAIN 598 701 CADHERIN_5.
FT DOMAIN 842 857 SER-RICH.
FT MOD_RES 842 842 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 844 844 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 850 850 PHOSPHORYLATION (BY SIMILARITY).
FT CARBOHYD 562 562 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 641 641 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 886 AA; 98714 MW; A9AE2BE8797A547 CRC64;

Query Match 1.8%; Score 15; DB 1; Length 886;
Best Local Similarity 100.0%; Pred. No. 1.3e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 744 YRPRPNDPEIGNFI 758
Db 801 YRPRPNDPEIGNFI 815

RESULT 13
CAD4_CHICK
ID CAD4_CHICK STANDARD; PRT; 913 AA.
AC P24503;
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cadherin-4 precursor (Retinal-cadherin) (R-cadherin) (R-CAD).
GN CDH4.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_TaxID=9031;
RN [1]_TaxID=9031;
RP SEQUENCE FROM N.A.
RC TISSUE=Retina;
RX MEDLINE=91299341; PubMed=1712604;
RA Inuzuka H., Miyatani S., Takeichi M.;
RT "R-cadherin: a novel Ca(2+)-dependent cell-cell adhesion molecule
expressed in the retina.";
RL Neuron 7:69-79(1991).
CC !- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
CC They preferentially interact with themselves in a homophilic
CC manner in connecting cells; cadherins may thus contribute to the
CC sorting of heterogeneous cell types. May play an important role in
CC retinal development.
CC !- SUBCELLULAR LOCATION: Type I membrane protein.
CC !- TISSUE SPECIFICITY: Embryonic brain and neuronal retina.
CC !- DEVELOPMENTAL STAGE: Detected only after some degree of neuronal
```

```
CC differentiation has taken place and persists at least up to the
CC newly hatched stage.
CC !- SIMILARITY: Contains 5 cadherin domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D14459; BAA03356.1; -
CC PIR; JH0424; ICHCR.
CC HSPF; P15116; INCI.
CC InterPro; IPR002126; Cadherin.
CC InterPro; IPR002333; Cadherin_C_term.
CC Pfam; PF00028; cadherin; 5.
CC Pfam; PF01049; Cadherin_C_term; 1.
CC PRINTS; PR00205; CADHERIN.
CC SMART; SM00112; CA; 5.
CC PROSITE; PS00232; CADHERIN_1; 3.
CC PROSITE; PS0268; CADHERIN_2; 5.
CC Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
KW Signal.
FT SIGNAL 1 19 POTENTIAL.
FT PROPEP 20 166 POTENTIAL.
FT CHAIN 167 913 CADHERIN-4.
FT DOMAIN 167 731 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 732 753 POTENTIAL.
FT DOMAIN 754 913 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 167 274 CADHERIN_1.
FT DOMAIN 275 389 CADHERIN_2.
FT DOMAIN 390 504 CADHERIN_3.
FT DOMAIN 505 610 CADHERIN_4.
FT DOMAIN 611 721 CADHERIN_5.
FT DOMAIN 870 885 SER-RICH.
FT CARBOHYD 280 280 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 409 409 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 554 554 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 629 629 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 658 658 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 699 699 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 652 652 P -> T (IN ONE FORM).
SQ SEQUENCE 913 AA; 100885 MW; BD2BB9ECC815DD6 CRC64;

Query Match 1.6%; Score 13; DB 1; Length 913;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 815 RFKKLADMYGGGE 827
Db 899 RFKKLADMYGGGE 911

RESULT 14
CAD4_MOUSE
ID CAD4_MOUSE STANDARD; PRT; 913 AA.
AC P39038;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Cadherin-4 precursor (Retinal-cadherin) (R-cadherin) (R-CAD).
GN CDH4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]_TaxID=10090;
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreatic islets;
RX MEDLINE=94067164; PubMed=8247017;
RA Hutton J.C., Christofori G., Chi W.Y., Edman U., Guest P.C.,
```

RA Hanahan D., Kelly R.B.;
RT "Molecular cloning of mouse pancreatic islet R-cadherin: differential
RT expression in endocrine and exocrine tissue.";
RL Mol. Endocrinol. 7:1151-1160(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94095672; PubMed=8270638;
RA Matsunami H., Miyatani S., Inoue T., Copeland N.G., Gilbert D.,
RA Jenkins N.A., Takeichi M.;
RT "Cell binding specificity of mouse R-cadherin and chromosomal mapping
RT of the gene.";
RL J. Cell Sci. 106:401-409(1993).
CC -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
CC They preferentially interact with themselves in a homophilic
CC manner in connecting cells; cadherins may thus contribute to the
CC sorting of heterogeneous cell types. May play an important role
CC in retinal development.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: Distributed widely in mouse tissues with high
CC levels present in brain, skeletal muscle and thymus.
CC -!- SIMILARITY: Contains 5 cadherin domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X69966; CAA49589.1; -;
DR EMBL; D14888; BAA03605.1; -;
DR PIR; A47543; A47543.
DR HSP; P15116; INCI.
DR MGD; MGI:99218; Cdh4.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR000233; Cadherin_C_term.
DR Pfam; PF00028; cadherin; 5.
DR Pfam; PF01049; Cadherin_C_term; 1.
DR PRINTS; PRO0205; CADHERIN.
DR SMART; SM00112; CA; 5.
DR PROSITE; PS00232; CADHERIN_1; 3.
DR PROSITE; PS0268; CADHERIN_2; 5.
DR Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
KW Signal.
FT SIGNAL 1 20 POTENTIAL.
FT PROPEP 21 166 POTENTIAL.
FT CHAIN 167 913 CADHERIN-4.
FT DOMAIN 167 731 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 732 753 POTENTIAL.
FT DOMAIN 754 913 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 167 274 CADHERIN_1.
FT DOMAIN 275 389 CADHERIN_2.
FT DOMAIN 390 504 CADHERIN_3.
FT DOMAIN 505 610 CADHERIN_4.
FT DOMAIN 611 721 CADHERIN_5.
FT DOMAIN 870 885 SER-RICH.
FT CARBOHYD 146 146 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 280 280 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 409 409 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 554 554 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 629 629 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 658 658 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 699 699 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 913 AA; 100030 MW; 1245A8CE8C338481 CRC64;
Query Match 1.6%; Score 13; DB 1; Length 913;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 815 RFKGLADWYGGGE 827
DB 899 RFKGLADWYGGGE 911

RESULT 15.
CAD4 HUMAN
ID CAD4 HUMAN STANDARD; PRT; 916 AA.
AC P55283;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cadherin-4 precursor (Retinal-cadherin) (R-cadherin) (R-CAD).
DE CDH4.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain.
RX MEDLINE=95073006; PubMed=7982033;
RA Tanihara H., Sano K., Helmark R.L., St John T., Suzuki S.;
RT "Cloning of five human cadherins clarifies characteristic features of
RT cadherin extracellular domain and provides further evidence for two
RT structurally different types of cadherin.";
RL Cell Adhes. Commun. 2:15-26(1994).
RN [2]
RP SEQUENCE OF 393-916 FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=91283540; PubMed=2059658;
RA Suzuki S., Sano K., Tanihara H.;
RT "Diversity of the cadherin family: evidence for eight new cadherins
RT in nervous tissue.";
RL Cell Regul. 2:261-270(1991).
CC -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
CC They preferentially interact with themselves in a homophilic
CC manner in connecting cells; cadherins may thus contribute to the
CC sorting of heterogeneous cell types. May play an important role
CC in retinal development.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: Expressed mainly in brain but also found in
CC other tissues.
CC -!- SIMILARITY: Contains 5 cadherin domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L34059; AAA35627.1; -;
DR PIR; C38992; C38992.
DR HSP; P15116; INCI.
DR Genew; HGNC:1763; CDH4.
DR MIM; 603006; -;
DR GO; GO:0007155; P:cell adhesion; TAS.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR000233; Cadherin_C_term.
DR Pfam; PF00028; cadherin; 5.
DR Pfam; PF01049; Cadherin_C_term; 1.
DR PRINTS; PRO0205; CADHERIN.
DR SMART; SM00112; CA; 5.
DR PROSITE; PS00232; CADHERIN_1; 2.
DR PROSITE; PS0268; CADHERIN_2; 5.
DR Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
KW Signal.
FT SIGNAL 1 20 POTENTIAL.
FT PROPEP 21 169 POTENTIAL.
FT CHAIN 170 916 CADHERIN-4.
FT DOMAIN 170 734 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 735 756 POTENTIAL.
FT DOMAIN 757 916 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 170 277 CADHERIN_1.

FT DOMAIN 278 392 CADHERIN 2.
FT DOMAIN 393 507 CADHERIN 3.
FT DOMAIN 508 613 CADHERIN 4.
FT DOMAIN 614 724 CADHERIN 5.
FT DOMAIN 873 888 SSR-RICH.
FT CARBOHYD 283 283 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 412 412 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 557 557 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 632 632 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 661 661 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 702 702 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 916 AA; 100446 MW; E717C54A19E0C52A CRC64;

Query Match 1.6%; Score 13; DB 1; Length 916;

Best Local Similarity 100.0%; Pred. No. 0.001;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 815 RFKKLADMYGGGE 827

Db 902 RFKKLADMYGGGE 914

Search completed: September 21, 2004, 22:25:56

Job time : 27 secs

This Page Blank (usp10)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 22:10:15 ; Search time 136 Seconds
(without alignments)
1923.269 Million cell updates/sec

Title: US-09-916-849A-1

Perfect score: 829

Sequence: 1 MGLPRGLASLLLLQVCWLQ.....NEWGSRFKKLADMYGGEDD 829

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL 25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriaph:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	51	6.2	753	11 Q80VY6	Q80VY6 mus musculus
2	51	6.2	821	11 Q83RE1	Q83RE1 mus musculus
3	51	6.2	822	11 Q83SL6	Q83SL6 mus musculus
4	37	4.5	105	11 Q9JIV6	Q9JIV6 rattus norv
5	15	1.8	821	4 Q9JII8	Q9JII8 homo sapien
6	15	1.8	901	4 Q9JII7	Q9JII7 homo sapien
7	14	1.7	40	6 Q18222	Q18222 sus scrofa
8	14	1.7	864	13 Q90Z37	Q90Z37 brachydanio
9	14	1.7	2909	5 Q8MY78	Q8MY78 asterina pe
10	13	1.6	566	4 Q9HIG6	Q9HIG6 homo sapien
11	13	1.6	798	5 Q9NLA7	Q9NLA7 ciona intes
12	13	1.6	824	4 Q8NB64	Q8NB64 homo sapien
13	12	1.4	68	11 Q80XX1	Q80XX1 mus sp. n-c
14	12	1.4	140	4 Q8ND09	Q8ND09 homo sapien
15	12	1.4	490	13 Q9W6G5	Q9W6G5 brachydanio
16	12	1.4	906	4 Q8N173	Q8N173 homo sapien

17	12	1.4	906	11 Q8BS19	Q8BS19 mus musculus
18	12	1.4	922	13 P79883	P79883 xenopus ma
19	11	1.3	111	11 Q9JIV7	Q9JIV7 rattus norv
20	11	1.3	792	13 Q9DFS1	Q9DFS1 xenopus lae
21	11	1.3	792	13 Q9DFS0	Q9DFS0 xenopus lae
22	10	1.2	66	6 Q8HX20	Q8HX20 bos taurus
23	10	1.2	177	12 Q8QVM3	Q8QVM3 hepatitis e
24	10	1.2	674	12 Q7TGF1	Q7TGF1 hepatitis e
25	10	1.2	1959	5 Q8MY75	Q8MY75 ptychodera
26	9	1.1	105	11 Q9JII1	Q9JII1 rattus norv
27	9	1.1	124	16 Q82ET9	Q82ET9 streptomyce
28	9	1.1	162	11 Q80Z61	Q80Z61 rattus norv
29	9	1.1	386	16 Q8RE76	Q8RE76 thermoanaer
30	9	1.1	393	16 Q8PVM2	Q8PVM2 xanthomonas
31	9	1.1	393	16 Q8PB63	Q8PB63 xanthomonas
32	9	1.1	418	13 Q7ZVY0	Q7ZVY0 brachydanio
33	9	1.1	481	16 Q92VS2	Q92VS2 rhizobium m
34	9	1.1	612	8 Q94TA9	Q94TA9 neoscorpelus
35	9	1.1	681	11 Q8BHP7	Q8BHP7 mus musculus
36	8	1.0	27	2 Q44952	Q44952 borrelia bu
37	8	1.0	34	4 Q96IX0	Q96IX0 homo sapien
38	8	1.0	40	16 Q9IIX6	Q9IIX6 pseudomonas
39	8	1.0	59	5 Q61591	Q61591 ostertagia
40	8	1.0	105	3 Q9P8L0	Q9P8L0 gibberella
41	8	1.0	106	11 Q9JII2	Q9JII2 rattus norv
42	8	1.0	120	5 P90945	P90945 caenorhabdi
43	8	1.0	132	16 Q9RXV0	Q9RXV0 deinococcus
44	8	1.0	136	10 Q7XQY7	Q7XQY7 cryza sativ
45	8	1.0	140	4 Q9NXX2	Q9NXX2 homo sapien

ALIGNMENTS

RESULT 1

Q80VY6	PRELIMINARY;	PRT;	753 AA.
ID	Q80VY6		
AC	Q80VY6		
DT	01-JUN-2003 (TRENBLrel. 24, Created)		
DT	01-JUN-2003 (TRENBLrel. 24, Last sequence update)		
DT	01-OCT-2003 (TRENBLrel. 25, Last annotation update)		
DE	Hypothetical protein (Fragment).		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Embryo;		
RX	MEDLINE=23388257; PubMed=12477932;		
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,		
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,		
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,		
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,		
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,		
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,		
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,		
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,		
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,		
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,		
RA	Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,		
RA	Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,		
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,		
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,		
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,		
RA	Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,		
RT	"Generation and initial analysis of more than 15,000 full-length human		
RT	and mouse cDNA sequences."		
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Embryo;		


```

RA Strausberg R.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC052189; AAF52189.1; -.
DR GO; GO:016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR00233; Cadherin_C_term.
DR Pfam; PF00028; cadherin; 5.
DR Pfam; PF01049; Cadherin_C_term; 1.
DR PRINTS; PR00205; CADHERIN.
DR SMART; SM00112; CA; 4.
DR PROSITE; PS00232; CADHERIN_1; 3.
DR PROSITE; PS0268; CADHERIN_2; 4.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 753 AA; 83300 MW; 414AFB4059CE94D2 CRC64;

Query Match 6.2%; Score 51; DB 11; Length 753;
Best Local Similarity 100.0%; Pred. No. 2.9e-43;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 684 KEPILLPEDTRDNVYFYGEGGEDDYDTQLHRGLEARPEVVLNDV 734
DB 608 KEPILLPEDTRDNVYFYGEGGEDDYDTQLHRGLEARPEVVLNDV 658

RESULT 2
Q8BSL6 PRELIMINARY; PRT; 821 AA.
AC Q8BSL6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cadherin 3.
GN CDH3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Forelimb;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium.
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK031265; BAC27327.1; -.
DR MGD; MGI:88356; Cdh3.
DR GO; GO:0005886; C:plasma membrane; IDA.
DR InterPro; IPR002126; Cadherin_1; 3.
DR InterPro; IPR00233; Cadherin_C_term.
DR Pfam; PF00028; cadherin; 5.
DR Pfam; PF01049; Cadherin_C_term; 1.
DR PRINTS; PR00205; CADHERIN.
DR SMART; SM00112; CA; 4.
DR PROSITE; PS00232; CADHERIN_1; 3.
DR PROSITE; PS0268; CADHERIN_2; 4.
SQ SEQUENCE 821 AA; 90512 MW; 169B56C5FB5CCBD0 CRC64;

Query Match 6.2%; Score 51; DB 11; Length 821;
Best Local Similarity 100.0%; Pred. No. 2.9e-43;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 684 KEPILLPEDTRDNVYFYGEGGEDDYDTQLHRGLEARPEVVLNDV 734
DB 676 KEPILLPEDTRDNVYFYGEGGEDDYDTQLHRGLEARPEVVLNDV 726

RESULT 3
Q8BSL6 PRELIMINARY; PRT; 822 AA.
AC Q8BSL6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cadherin 3.
GN CDH3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Forelimb;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium.
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK045041; BAC32194.1; -.
DR MGD; MGI:88356; Cdh3.
DR GO; GO:0005886; C:plasma membrane; IDA.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR00233; Cadherin_C_term.
DR Pfam; PF00028; cadherin; 5.
DR Pfam; PF01049; Cadherin_C_term; 1.
DR PRINTS; PR00205; CADHERIN.
DR SMART; SM00112; CA; 4.
DR PROSITE; PS00232; CADHERIN_1; 3.
DR PROSITE; PS0268; CADHERIN_2; 4.
SQ SEQUENCE 822 AA; 90611 MW; 7C653D58210A595A CRC64;

Query Match 6.2%; Score 51; DB 11; Length 822;
Best Local Similarity 100.0%; Pred. No. 2.9e-43;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 684 KEPILLPEDTRDNVYFYGEGGEDDYDTQLHRGLEARPEVVLNDV 734
DB 677 KEPILLPEDTRDNVYFYGEGGEDDYDTQLHRGLEARPEVVLNDV 727

RESULT 4
Q8BSL6 PRELIMINARY; PRT; 105 AA.
AC Q8BSL6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE P-cadherin (Fragment).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fischer344; TISSUE=Testis;
RX MEDLINE=20114371; PubMed=10650949;
RA Johnson K.J., Patel S.R., Boskelheide K.;
RT "Multiple cadherin superfamily members with unique expression profiles
RT are produced in rat testis."
RL Endocrinology 141:675-683(2000).
DR EMBL; AF177683; AAF87058.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
DR InterPro; IPR00233; Cadherin_C_term.
DR Pfam; PF01049; Cadherin_C_term; 1.
FT NON_TER 1
FT NON_TER 105
SQ SEQUENCE 105 AA; 11650 MW; 6F23CB9F66F1B27B CRC64;

Query Match 4.5%; Score 37; DB 11; Length 105;

```

```

Best Local Similarity 100.0%; Pred. No. 1.5e-29; Indels 0; Gaps 0;
Matches 37; Conservative 0; Mismatches 0;

QY 739 ITPMYRPANPDEIGNFIENLKAANTDPTAPPYD 775
    |||||
DB 29 ITPMYRPANPDEIGNFIENLKAANTDPTAPPYD 65
    |||||

RESULT 5
Q9UII8 PRELIMINARY; PRT; 821 AA.
AC Q9UII8;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE E-cadherin.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Shibamoto S., Fukudome Y., Yanagihara K.;
RT "Mutant E-cadherin."
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC EMBL; AB025105; BAA88956.1; -.
DR HSSP; P09803; 1SUH.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0005194; F:cell adhesion molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
DR InterPro; IPR002126; Cadherin.
DR Pfam; PF00028; cadherin; 4.
DR PRINTS; PR00205; CADHERIN.
DR SMART; SM00112; CA; 3.
DR PROSITE; PS00232; CADHERIN_1; 3.
DR PROSITE; PS0268; CADHERIN_2; 3.
DR Calcium; Calcium-binding; Cell adhesion; Glycoprotein.
KW SEQUENCE 821 AA; 90941 MW; 4EFT06672ACE35E9 CRC64;
SQ

Query Match 1.8%; Score 15; DB 4; Length 821;
Best Local Similarity 100.0%; Pred. No. 6.7e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 815 RFKLLADMYGGGDD 829
    |||||
DB 807 RFKLLADMYGGGDD 821
    |||||

RESULT 6
Q9UII7 PRELIMINARY; PRT; 901 AA.
AC Q9UII7;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE E-cadherin.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Shibamoto S., Fukudome Y.;
RT "E-cadherin mutant."
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC EMBL; AB025105; BAA88956.1; -.
DR HSSP; P09803; 1SUH.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0005194; F:cell adhesion molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
DR InterPro; IPR002126; Cadherin.
DR Pfam; PF00028; cadherin; 4.
DR PRINTS; PR00205; CADHERIN.
DR SMART; SM00112; CA; 3.
DR PROSITE; PS00232; CADHERIN_1; 3.
DR PROSITE; PS0268; CADHERIN_2; 3.
DR Calcium; Calcium-binding; Cell adhesion; Glycoprotein.
KW SEQUENCE 901 AA; 99693 MW; 42160D749BCF120E CRC64;
SQ

Query Match 1.8%; Score 15; DB 4; Length 901;
Best Local Similarity 100.0%; Pred. No. 7.3e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 815 RFKLLADMYGGGDD 829
    |||||
DB 887 RFKLLADMYGGGDD 901
    |||||

RESULT 7
O18922 PRELIMINARY; PRT; 40 AA.
AC O18922;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Epithelial-cadherin (E-cadherin) (Uvomorulin) (CaM 120/80) (Fragment).
GN CADL.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=RETINA;
RA Lutz D.A., Zheng J.J.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES. E-CADHERIN HAS A POTENT
CC INVASIVE SUPPRESSOR ROLE.
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
DR EMBL; AF033019; AAB87474.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0005194; F:cell adhesion molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
DR InterPro; IPR002126; Cadherin.
DR PRINTS; PR00205; CADHERIN.
DR SMART; SM00112; CA; 4.
DR PROSITE; PS00232; CADHERIN_1; PARTIAL.
DR PROSITE; PS0268; CADHERIN_2; PARTIAL.
KW Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;
KW Calcium-binding; Repeat.
FT NON_TER 1
FT CHAIN <1 >40 EPIITHELIAL-CADHERIN.

```

```

RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC EMBL; AB025106; BAA88957.1; -.
DR HSSP; P09803; 1SUH.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0005194; F:cell adhesion molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR002233; Cadherin_C_term.
DR Pfam; PF00028; cadherin; 5.
DR PRINTS; PR00205; CADHERIN.
DR SMART; SM00112; CA; 4.
DR PROSITE; PS00232; CADHERIN_1; 3.
DR PROSITE; PS0268; CADHERIN_2; 3.
DR Calcium; Calcium-binding; Cell adhesion; Glycoprotein.
KW SEQUENCE 901 AA; 99693 MW; 42160D749BCF120E CRC64;
SQ

Query Match 1.8%; Score 15; DB 4; Length 901;
Best Local Similarity 100.0%; Pred. No. 7.3e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 815 RFKLLADMYGGGDD 829
    |||||
DB 887 RFKLLADMYGGGDD 901
    |||||

RESULT 7
O18922 PRELIMINARY; PRT; 40 AA.
AC O18922;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Epithelial-cadherin (E-cadherin) (Uvomorulin) (CaM 120/80) (Fragment).
GN CADL.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=RETINA;
RA Lutz D.A., Zheng J.J.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES. E-CADHERIN HAS A POTENT
CC INVASIVE SUPPRESSOR ROLE.
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
DR EMBL; AF033019; AAB87474.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0005194; F:cell adhesion molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
DR InterPro; IPR002126; Cadherin.
DR PRINTS; PR00205; CADHERIN.
DR SMART; SM00112; CA; 4.
DR PROSITE; PS00232; CADHERIN_1; PARTIAL.
DR PROSITE; PS0268; CADHERIN_2; PARTIAL.
KW Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;
KW Calcium-binding; Repeat.
FT NON_TER 1
FT CHAIN <1 >40 EPIITHELIAL-CADHERIN.

```

FT DOMAIN <1 >40 CYTOPLASMIC (POTENTIAL).
 FT NON TER 40 40
 SQ SEQUENCE 40 AA; 4449 MW; CB928720A51B8372 CRC64;

Query Match 1.7%; Score 14; DB 6; Length 40;
 Best Local Similarity 100.0%; Pred. No. 4.5e-06;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 800 SDQDDQDYLYNNGW 813
 Db 25 SDQDDQDYLYNNGW 38

RESULT 8

Q90237 PRELIMINARY; PRT; 864 AA.
 AC Q90237;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE E-cadherin.
 OS Brachydanio rerio (zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OC NCBI_TaxID=7955;
 RN [1]
 RX MEDLINE=21270060; PubMed=11376490;
 RA Babb S.G., Barnett J., Doedens A.L., Cobb N., Liu Q., Sorkin B.C.,
 RA Velick P.C., Raymond P.A., Marrs J.A.;
 RT "Zebrafish E-cadherin: expression during early embryogenesis and
 RT regulation during brain development.";
 RL Dev. Dyn. 221:231-237(2001).
 CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
 CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
 CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
 CC SORTING OF HETEROGENEOUS CELL TYPES (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
 DE EMBL; AF364811; AAK52054.1; .
 DR ZFIN; ZDB-GENE-010606-1; cdh1.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005509; F:calcium ion binding; IEA.
 DR GO; GO:0005194; F:cell adhesion molecule activity; IEA.
 DR GO; GO:0007155; P:cell adhesion; IEA.
 DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
 DR InterPro; IPR002126; Cadherin.
 DR InterPro; IPR000233; Cadherin_C_term.
 DR Pfam; PF01049; Cadherin_C_term; 5.
 DR PRINTS; PR00205; CADHERIN.
 DR SMART; SM00112; CA; 4.
 DR PROSITE; PS00232; CADHERIN_1; 3.
 DR PROSITE; PS02288; CADHERIN_2; 4.
 DR Calcium; Calcium-binding; Cell adhesion; Glycoprotein.
 SQ SEQUENCE 864 AA; 95254 MW; 5ACA19538396397C CRC64;

Query Match 1.7%; Score 14; DB 13; Length 864;
 Best Local Similarity 100.0%; Pred. No. 7.6e-05;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 815 RFKKLADMYGGGED 828
 Db 851 RFKKLADMYGGGED 864

RESULT 9

Q8MY78 PRELIMINARY; PRT; 2909 AA.
 AC Q8MY78;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Ap-cadherin.
 GN APCAD.
 OS Asterina pectinifera (Starfish).
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Asterozoa;
 OC Asteroidea; Valvatacea; Valvatida; Asterinidae; Asterina.
 OC NCBI_TaxID=7594;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Oda H., Wada H., Tagawa K., Akiyama-Oda Y., Satoh N., Humphreys T.,
 RA Zhang S., Tsukita S.;
 RT "A novel amphioxus cadherin that localizes to epithelial adherens
 RT junctions has an unusual domain organization with implications for
 RT chordate phylogeny.";
 RL Evol. Dev. 0:0-0(2002).
 CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
 CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
 CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
 CC SORTING OF HETEROGENEOUS CELL TYPES (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
 DE EMBL; AB075365; BAC06934.1; .
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005509; F:calcium ion binding; IEA.
 DR GO; GO:0005194; F:cell adhesion molecule activity; IEA.
 DR GO; GO:0007155; P:cell adhesion; IEA.
 DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
 DR InterPro; IPR000152; Asx_Hydroxyl_S.
 DR InterPro; IPR002126; Cadherin.
 DR InterPro; IPR000233; Cadherin_C_term.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF_Like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR001791; Laminin_G.
 DR Pfam; PF00028; cadherin; 18.
 DR Pfam; PF01049; Cadherin_C_term; 1.
 DR Pfam; PF00008; EGF; 4.
 DR Pfam; PF00054; laminin_G; 2.
 DR PRINTS; PR00205; CADHERIN.
 DR SMART; SM00112; CA; 17.
 DR SMART; SM00181; EGF; 4.
 DR SMART; SM00179; EGF_CA; 4.
 DR SMART; SM00282; LamG; 2.
 DR PROSITE; PS00010; ASX HYDROXYL; 2.
 DR PROSITE; PS00232; CADHERIN_1; 5.
 DR PROSITE; PS02268; CADHERIN_2; 17.
 DR PROSITE; PS00022; EGF_1; 3.
 DR PROSITE; PS01186; EGF_2; 4.
 DR PROSITE; PS00025; LAM_G_DOMAIN; 2.
 DR Calcium; Calcium-binding; Cell adhesion; EGF-like domain;
 KW Glycoprotein.
 SQ SEQUENCE 2909 AA; 316542 MW; 2466441EC36619CF CRC64;

Query Match 1.7%; Score 14; DB 5; Length 2909;
 Best Local Similarity 100.0%; Pred. No. 0.00023;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 816 FKKLADMYGGGED 829
 Db 2896 FKKLADMYGGGED 2909

RESULT 10

Q9H1G6 PRELIMINARY; PRT; 566 AA.
 ID Q9H1G6
 AC Q9H1G6;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE DULL07C24.2 (Continued from BA723K16.1 in Em:AL365229, BA429B11.1 in
 DE Em:AL162457 and BA489M19.1 in Em:AL365401) (Fragment).
 GN CH4.
 OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wall M.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; AL109911; CAC22311.1; -.
DR HSSP; P15116; INCU.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0005194; F:cell adhesion molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
DR InterPro; IPR002126; Cadherin_C_term.
DR InterPro; IPR000233; Cadherin_C_term.
DR Pfam; PF01049; Cadherin_C_term; 1.
DR PRINTS; PR00205; CADHERIN.
DR SMART; SM00112; CA; 3.
DR PROSITE; PS00232; CADHERIN_1; 2.
DR PROSITE; PS0268; CADHERIN_2; 2.
DR Calcium; Calcium-binding; Cell adhesion; Glycoprotein.
KW NON TER
FT 1
SQ SEQUENCE 566 AA; 61628 MW; 99A9ADD91B5AE7A3 CRC64;

Query Match 1.6%; Score 13; DB 4; Length 566;
Best Local Similarity 100.0%; Pred. No. 0.00056;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 815 RPKKLADMYGGGE 827
DB 552 RPKKLADMYGGGE 564
|||||

RESULT 11
Q9NL47 PRELIMINARY; PRT; 798 AA.
ID Q9NL47
AC Q9NL47
DT 01-OCT-2000 (T-EMBLrel. 15, Created)
DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Cadherin.
GN C1-CADHERIN.
OS Ciona intestinalis.
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
OC Phlebobranchia; Cionidae; Ciona.
OX NCBI_TaxID=7719;
RN [1]
RP SEQUENCE FROM N.A.
RA Imai K., Takada N., Satoh N., Satou Y.;
RT "An essential role of beta-catenin in the endoderm specification of
RT ascidian embryo."
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; AB031540; BAA92182.1; -.
DR HSSP; P15116; INCU.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0005194; F:cell adhesion molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR000233; Cadherin_C_term.

DR Pfam; PF00028; cadherin; 4.
DR PRINTS; PR00205; CADHERIN.
DR SMART; SM00112; CA; 5.
DR PROSITE; PS00232; CADHERIN_1; 3.
DR PROSITE; PS0268; CADHERIN_2; 5.
KW Calcium; Calcium-binding; Cell adhesion; Glycoprotein.
SQ SEQUENCE 798 AA; 86140 MW; 6F13E42D9F230A5C CRC64;

Query Match 1.6%; Score 13; DB 5; Length 798;
Best Local Similarity 100.0%; Pred. No. 0.00077;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 815 RPKKLADMYGGGE 827
DB 784 RPKKLADMYGGGE 796
|||||

RESULT 12
Q9NB64 PRELIMINARY; PRT; 824 AA.
ID Q9NB64
AC Q9NB64
DT 01-OCT-2002 (T-EMBLrel. 22, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ34177.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,
RA Hotuta T., Hiracka S., Murakawa K., Takiguchi S., Kusano J., Chiba Y.,
RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
RA Sugiyama T., Irie R., Otsuki T., Sato H., Ota T., Wakamatsu A.,
RA Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wacatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT "NED0 human cDNA sequencing project."
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; AK031456; BAC03677.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0005194; F:cell adhesion molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR000233; Cadherin_C_term.
DR Pfam; PF00028; cadherin; 5.
DR Pfam; PF01049; Cadherin_C_term; 1.
DR PRINTS; PR00205; CADHERIN.
DR SMART; SM00112; CA; 5.
DR PROSITE; PS00232; CADHERIN_1; 3.
DR PROSITE; PS0268; CADHERIN_2; 3.
KW Hypothetical protein; Calcium; Calcium-binding; Cell adhesion;
KW Glycoprotein.
SQ SEQUENCE 824 AA; 90261 MW; 31A0C127BD40BAEB CRC64;

Query Match 1.6%; Score 13; DB 4; Length 824;
Best Local Similarity 100.0%; Pred. No. 0.0008;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 815 RPKKLADMYGGGE 827
DB 810 RPKKLADMYGGGE 822
|||||

```

RESULT 13
Q80XX1 PRELIMINARY; PRT; 68 AA.
AC Q80XX1; 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE N-cadherin (Fragment).
GN N-CADHERIN.
OS Mus sp.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10095;
[1]
RN RP SEQUENCE FROM N.A.
RX MEDLINE=92409532; PubMed=1528849;
RA Miyatani S., Copeland N.G., Gilbert D.J., Jenkins N.A., Takeichi M.;
RT "Genomic structure and chromosomal mapping of the mouse N-cadherin
RT Gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:8443-8447(1992).
DR EMBL; S45013; AAP13881.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
DR InterPro; IPR000233; Cadherin_C_term.
DR Pfam; PF01049; Cadherin_C_term; 1.
FT NON_TER 1
SQ SEQUENCE 68 AA; 7130 MW; 860927ADDC8D6595 CRC64;

Query Match 1.4%; Score 12; DB 11; Length 68;
Best Local Similarity 100.0%; Pred. No. 0.0087;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 815 RFKKLADMYGGG 826
DB 55 RFKKLADMYGGG 66

RESULT 14
Q8ND09 PRELIMINARY; PRT; 140 AA.
AC Q8ND09;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein.
GN DRFP434K1130.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN RP SEQUENCE FROM N.A.
RX TISSUE=Testis;
RA Blum H., Bauersachs S., Mewes H.W., Weil B., Wiemann S.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL834206; CAD38893.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
DR InterPro; IPR000233; Cadherin_C_term.
DR Pfam; PF01049; Cadherin_C_term; 1.
KW Hypothetical protein.
SQ SEQUENCE 140 AA; 15832 MW; C4B81EBFB31F812B CRC64;

Query Match 1.4%; Score 12; DB 4; Length 140;
Best Local Similarity 100.0%; Pred. No. 0.0017;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 703 EGGGEEDQDYD 714
DB 359 EGGGEEDQDYD 370

Query Match 1.4%; Score 12; DB 13; Length 490;
Best Local Similarity 100.0%; Pred. No. 0.0054;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 703 EGGGEEDQDYD 714
DB 359 EGGGEEDQDYD 370

Search completed: September 21, 2004, 22:25:24
Job time : 138 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 22:09:44 ; Search time 132 Seconds
(without alignments)
1774.484 Million cell updates/sec

Title: US-09-916-849a-1

Perfect score: 829

Sequence: 1 MGLPRGPLASLLILQVWLQ.....NEWGSRFKKLADMYGGGDD 829

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	829	100.0	829	5	ABG61897
2	829	100.0	829	5	ABG61897 Prostate
3	829	100.0	829	5	Abb81476 Human P-C
4	829	100.0	829	5	Abj05598 Breast ca
5	829	100.0	829	5	Aau97432 Human P-C
6	829	100.0	829	5	Aam50864 Cadherin
7	829	100.0	829	6	Abp58357 Human P-C
8	829	100.0	829	7	Adc15497 Human bas
9	829	100.0	829	7	Adp14190 Human src
10	728	87.8	829	5	Abp54683 Metastati
11	728	87.8	829	6	Abp58670 Human can
12	728	87.8	829	6	Abu56434 Lung can
13	673	81.2	829	6	Abu56670 Lung can
14	108	13.0	108	2	Aay09093 Human P-C
15	108	13.0	108	2	Aay17104 Human P-C
16	108	13.0	108	3	Aay78184 Human P-C
17	108	13.0	108	4	Aag65364 Human P-C
18	108	13.0	108	5	Aam47471 Human P-C
19	108	13.0	108	5	Abu50282 Human P-C
20	108	13.0	108	7	Abu01245 Human cad
21	106	12.8	106	3	Aay78193 Human P-C
22	106	12.8	106	6	Abu60291 Human P-C
23	106	12.8	106	7	Abw01254 Human cad
24	48	5.8	78	2	Aay08239 Human cad
25	43	5.2	43	2	Aau91055 Flanking

26 43 5.2 43 3 AAB27069 Beta-cate
27 22 2.7 108 2 AAY09094 Mouse P-C
28 22 2.7 108 2 AAY17105 Mouse P-C
29 22 2.7 108 3 AAY78185 Mouse P-C
30 22 2.7 108 4 AAG65365 Mouse P-C
31 22 2.7 108 5 AAM47472 Murine P-
32 22 2.7 108 6 ABU60283 Mouse P-C
33 22 2.7 108 7 ABO43534 Mouse P-C
34 22 2.7 108 7 ABW01246 Mouse cad
35 19 2.3 19 5 AAM50867 Cadherin
36 19 2.3 19 7 ADC15500 Human bas
37 17 2.1 17 5 AAM50868 Cadherin
38 17 2.1 17 7 ADC15501 Human bas
39 15 1.8 15 5 AAM50869 Cadherin
40 15 1.8 15 7 ADC15502 Human bas
41 15 1.8 882 2 AAY09375 Wild-type
42 15 1.8 882 3 AAB35730 Human E-C
43 15 1.8 882 4 AAB73490 Human E-C
44 15 1.8 882 5 ABG96293 Human ova
45 15 1.8 882 5 AAU78051 Human E-C

ALIGNMENTS

RESULT 1

ABG61897
ID ABG61897 standard; protein; 829 AA.

XX ABG61897;

DT 15-AUG-2002 (first entry)

DE Prostate cancer-associated protein #98.

KW Prostate cancer; prostate tumour tissue; human; mammal; cytostatic.

OS Mammalia.

PN WO200230268-A2.

XX 18-APR-2002.

XX 12-OCT-2001; 2001WO-US032045.

PR 13-OCT-2000; 2000US-00687576.

PR 08-DEC-2000; 2000US-0073288.

PR 08-DEC-2000; 2000US-0073342.

PR 24-JAN-2001; 2001US-0263957P.

PR 16-MAR-2001; 2001US-0276791P.

PR 16-MAR-2001; 2001US-0276888P.

PR 06-APR-2001; 2001US-0281922P.

PR 24-APR-2001; 2001US-0286214P.

PR 30-APR-2001; 2001US-00847046.

PR 04-MAY-2001; 2001US-0288589P.

XX (EOSB-) EOS BIOTECHNOLOGY INC.

XX Gish KC, Mack DH, Wilson KE, Afar D, Hevezi P;

XX WPI, 2002-471335/50.

XX N-PSDB; ABK92214.

XX Detecting a prostate cancer-associated transcript in a cell in a patient,

XX useful for diagnosing prostate cancer (PC) or screening modulators of PC,

XX by determining if prostate cancer-associated genes are expressed in a

XX prostate tissue.

XX Claim 27; Page 384-385; 436pp; English.

XX The present invention relates to methods of detecting a prostate cancer-

XX associated transcript in a cell from a patient. The method comprises

XX contacting a biological sample from the patient with prostate cancer-

CC associated polynucleotides (designated PC genes) that selectively
 CC hybridise to a sequence that is at least 80% identical to them. The
 CC prostate cancer-associated polynucleotide sequences are differentially
 CC expressed in prostate tumour tissue or in prostate cancer and are derived
 CC from the tissues of various organisms such as humans or other mammals
 CC (e.g. mice, sheep and dogs). The methods of the invention are useful for
 CC diagnosing and treating prostate cancer in mammals. The prostate cancer-
 CC associated genes are useful for diagnosing or treating prostate cancer,
 CC as well as for identifying modulators of prostate cancer or agents that
 CC inhibit prostate cancer. The nucleic acid sequences are particularly
 CC useful in gene therapy, as a vaccine or in antisense applications.
 CC ABG61800-ABG61944 represent prostate cancer-associated proteins
 XX
 XX Sequence 829 AA;

Query Match 100.0%; Score 829; DB 5; Length 829;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPRGPLASLLQLQVLCQAASEPCRAVFAEVTLEAGGAEPGQALGVFMGCPG 60
 DB 1 MGLPRGPLASLLQLQVLCQAASEPCRAVFAEVTLEAGGAEPGQALGVFMGCPG 60

QY 61 QEPALFSTNDNDFTVRNGETVQERRSLKERNPLKIPFSKRILRRHKRDVWVAPISVPENG 120
 DB 61 QEPALFSTNDNDFTVRNGETVQERRSLKERNPLKIPFSKRILRRHKRDVWVAPISVPENG 120

QY 121 KGPPFQRLNCKNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWLANKLPDREETA 180
 DB 121 KGPPFQRLNCKNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWLANKLPDREETA 180

QY 181 YELFGHAUSENGASVEDPMNISIIVTDQNDHKPKFTQDTRFGSVLEGVLPGTSVMQVAT 240
 DB 181 YELFGHAUSENGASVEDPMNISIIVTDQNDHKPKFTQDTRFGSVLEGVLPGTSVMQVAT 240

QY 241 DEDDAITYNGVAVYATHQEPKDPHDMFTTHRSSTISVSSGLDREKVPYTLTQA 300
 DB 241 DEDDAITYNGVAVYATHQEPKDPHDMFTTHRSSTISVSSGLDREKVPYTLTQA 300

QY 301 TMDGSGSTTTAVAVVEILDANDNAPMFPDPQKYEAVHPENAVGHEVQRLTVTDLPNSP 360
 DB 301 TMDGSGSTTTAVAVVEILDANDNAPMFPDPQKYEAVHPENAVGHEVQRLTVTDLPNSP 360

QY 361 AWRATYLMGDDGDHFTTHPESNQGLTTRKGLDPEAKNQHTLYVEVTNEAPFVLK 420
 DB 361 AWRATYLMGDDGDHFTTHPESNQGLTTRKGLDPEAKNQHTLYVEVTNEAPFVLK 420

QY 421 PTSTATIIVHVEDVNEAPVFPVPSKVVEVQBGIPGEPVCVVTABDPKKNOKISYRI 480
 DB 421 PTSTATIIVHVEDVNEAPVFPVPSKVVEVQBGIPGEPVCVVTABDPKKNOKISYRI 480

QY 481 DPAGWLAMPDSQVAVGLDREDQFVRNNIYEVWVLAMDNQSPPTGTGTLTLTLD 540
 DB 481 DPAGWLAMPDSQVAVGLDREDQFVRNNIYEVWVLAMDNQSPPTGTGTLTLTLD 540

QY 541 VNDHGVPPEPQITICNQSPVRHLNITDKLSPTSPFQALTDSDIYTWAEVNEGSD 600
 DB 541 VNDHGVPPEPQITICNQSPVRHLNITDKLSPTSPFQALTDSDIYTWAEVNEGSD 600

QY 601 TVVLSLKLKQDPTDVHLSLSDHGNKEQLTVIRATVCDCHGVETCPGPKGFIPLVL 660
 DB 601 TVVLSLKLKQDPTDVHLSLSDHGNKEQLTVIRATVCDCHGVETCPGPKGFIPLVL 660

QY 661 GAVLALLFLLVLLLVKRRKIKPELLLPEDTDRDNVYEGEGGEDDQDIDTQLHR 720
 DB 661 GAVLALLFLLVLLLVKRRKIKPELLLPEDTDRDNVYEGEGGEDDQDIDTQLHR 720

QY 721 GLEAREVVLRLNDVAPTIPTPMYRPRANPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
 DB 721 GLEAREVVLRLNDVAPTIPTPMYRPRANPDEIGNFIENLKAANTDPTAPPYDTLLVF 780

QY 781 DYEGSGSDAASLSLTSSASDQDQDYDLNWSGRFVKKLADMYGGEDD 829

DB 781 DYEGSGSDAASLSLTSSASDQDQDYDLNWSGRFVKKLADMYGGEDD 829

RESULT 2
 ABB81476
 ID ABB81476 standard; protein; 829 AA.
 AC ABB81476;
 XX
 XX 30-AUG-2002 (first entry)
 DT
 XX Human P-cadherin protein SEQ ID NO:50.
 DE
 XX Cadherin; alpha-catenin; cancer; beta-catenin binding domain; melanoma;
 KW o-catenin; colon cancer.
 XX Homo sapiens.
 OS
 XX US2002045591-A1.
 PN
 XX 18-APR-2002.
 PD
 XX 17-JUL-2001; 2001US-00905983.
 PF
 XX 26-MAY-1998; 98IL-00124650.
 PR
 XX 26-MAY-1999; 99US-00318633.
 PP
 XX (GEIG/) GEIGER B.
 PA (BENZ/) BEN-ZE'EV A.
 PA (SADOT/) SADOT E.
 XX
 XX Geiger B, Ben-Ze'ev A, Sadot E;
 PI
 XX WPI; 2002-499105/53.
 DR N-PSDB; ABN83393.
 XX
 XX New construct encoding soluble cytoplasmic portion of cadherin including
 PT beta catenin binding domain useful in treating cancer associated with
 PT high beta-catenin activity e.g. colon cancer and melanoma.
 XX
 XX Example 3; Page 53-55; 102pp; English.

CC The present invention describes a pharmaceutical composition for treating
 CC cancer associated with abnormally high beta-catenin activity. The
 CC pharmaceutical composition comprises a gene therapy vehicle harbouring a
 CC polynucleotide that contains: (i) a nucleotide sequence encoding a
 CC soluble cytoplasmic portion of a cadherin which lacks a transmembrane
 CC portion and an extracellular portion of the cadherin, and includes a beta
 CC -catenin binding domain; and (b) an upstream promoter for directing
 CC expression of the soluble cytoplasmic portion of the cadherin in a
 CC mammalian cell. Also described is a pharmaceutical composition for
 CC treating cancer associated with abnormally high activity levels of beta-
 CC catenin comprising a nucleotide sequence encoding an o-catenin; and (b)
 CC that contains: (a) a nucleotide sequence encoding an o-catenin; and (b)
 CC an upstream promoter for directing expression of the o-catenin in a
 CC mammalian cell. The pharmaceutical compositions have cytostatic activity
 CC and can be used in the suppression of beta-catenin-mediated
 CC transactivation. They can be used for treating cancers associated with
 CC abnormally high activity levels of beta-catenin such as colon cancers and
 CC melanomas, by reducing these high activity levels of beta-catenin in
 CC mammalian cells. The present sequence represents human P-cadherin which
 CC is used in the exemplification of the present invention

XX
 SQ Sequence 829 AA;

Query Match 100.0%; Score 829; DB 5; Length 829;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPRGPLASLLQLQVLCQAASEPCRAVFAEVTLEAGGAEPGQALGVFMGCPG 60
 DB 1 MGLPRGPLASLLQLQVLCQAASEPCRAVFAEVTLEAGGAEPGQALGVFMGCPG 60

QY 421 PTSTATIVHVEDVNEAPVFPBPSKVVEVQSGIPTGEPVCVYTAEDPDKENQKISYRIILR 480
 DB 421 PTSTATIVHVEDVNEAPVFPBPSKVVEVQSGIPTGEPVCVYTAEDPDKENQKISYRIILR 480
 QY 481 DPAGWLAMDPSGQVTAAGTGLDREDEQFVRNNIYEVNVLAMONGSPPTGTGTLTLTLLD 540
 DB 481 DPAGWLAMDPSGQVTAAGTGLDREDEQFVRNNIYEVNVLAMONGSPPTGTGTLTLTLLD 540
 QY 541 VNHGVPPEPRQITICNQSPVRHLNITDKLSPHTSPFOAQLTDDSDIYWTAEVNEEGD 600
 DB 541 VNHGVPPEPRQITICNQSPVRHLNITDKLSPHTSPFOAQLTDDSDIYWTAEVNEEGD 600
 QY 601 TVVLSLKKFLKQDITYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFLLPVL 660
 DB 601 TVVLSLKKFLKQDITYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFLLPVL 660
 QY 661 GAVLALLFLLLVLLVLLVRKKRKIKEPILLPEDDTRDNVFFYBEGGGEEDQDYDITQLHR 720
 DB 661 GAVLALLFLLLVLLVLLVRKKRKIKEPILLPEDDTRDNVFFYBEGGGEEDQDYDITQLHR 720
 QY 721 GLEAPBEVLRNDVAPTIIPTPMYRPRPANDPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
 DB 721 GLEAPBEVLRNDVAPTIIPTPMYRPRPANDPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
 QY 781 DYEGSGSDAASLSLTSSASDQDQDYDYLNEWGSRFKKLADMYGGGEDD 829
 DB 781 DYEGSGSDAASLSLTSSASDQDQDYDYLNEWGSRFKKLADMYGGGEDD 829
 RESULT 4
 AAU97492
 ID AAU97492 standard; protein; 829 AA.
 AC AAU97492;
 XX
 DT 13-AUG-2002 (first entry)
 DE Human P-cadherin (placental cadherin) protein.
 KW Human; P-cadherin; placental cadherin; solid surface; vascular graft;
 KW cell-to-cell cohesion; endothelial cell-coated surface; vascular
 KW hydrodynamic shear; shaking container; continuous harvest system;
 KW laminar shear device.
 XX
 OS Homo sapiens.
 XX
 PN WO200231121-A2.
 XX
 PD 18-APR-2002.
 XX
 PF 15-OCT-2001; 2001WO-US032030.
 XX
 PR 13-OCT-2000; 2000US-0241216P.
 PR 27-OCT-2000; 2000US-0243693P.
 PR 11-OCT-2001; 2001US-00975723.
 XX
 PA (UNYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.
 XX
 PI Nackman G, Foty RA;
 XX
 DR WPI; 2002-435443/46.
 DR N-PSDB; ABK52038.
 XX
 PT Populating solid surface especially biomedical devices with vascular
 PT endothelial cells by increasing cell-to-cell cohesion useful to heal
 FT vascular grafts.
 XX
 PS Disclosure; Page 9-10; 38pp; English.
 XX
 CC The present invention relates to a new method of populating a solid
 CC surface with cells. The method of the invention involves increasing the
 CC cell-to-cell cohesion of the cells. The method is useful for populating a
 CC solid surface with human and nonhuman cells. The invention is also useful

CC for achieving better utilisation of vascular grafts and also in any
 CC system that employs endothelial cell-coated surface susceptible to
 CC hydrodynamic shear, such as shaking container coated with endothelial
 CC cells, continuous harvest systems for harvesting the products of
 CC endothelial cells on solid surface, and laminar shear devices. The
 CC present amino acid sequence represents the human P-cadherin (placental
 CC cadherin) protein of the invention
 XX
 SQ Sequence 829 AA;
 Query Match 100.0%; Score 829; DB 5; Length 829;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MGLPRGLASLLILLOVLCWLOCAASEPCRAVFAEVTLEAGGAEOBQALGKVFVGCPC 60
 DB 1 MGLPRGLASLLILLOVLCWLOCAASEPCRAVFAEVTLEAGGAEOBQALGKVFVGCPC 60
 QY 61 QEPALFSTDDNDDFTVRNGETVQERRSLKERNPLKIPPSKILRRHKEDWVVAISPENG 120
 DB 61 QEPALFSTDDNDDFTVRNGETVQERRSLKERNPLKIPPSKILRRHKEDWVVAISPENG 120
 QY 121 KGPPQRLNQLKSKNDRTXIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDRBIAK 180
 DB 121 KGPPQRLNQLKSKNDRTXIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDRBIAK 180
 QY 181 YELFGHAVSENGASVEDPMNISIIVTDNDHKPKFTQDTFRGSLVGLGFTSVMOVTAT 240
 DB 181 YELFGHAVSENGASVEDPMNISIIVTDNDHKPKFTQDTFRGSLVGLGFTSVMOVTAT 240
 QY 241 DEDDAIYTYNGWVAYSIHSEPKDPLMFTIHRSTGTISVISGLDRKPYEYTLTIOA 300
 DB 241 DEDDAIYTYNGWVAYSIHSEPKDPLMFTIHRSTGTISVISGLDRKPYEYTLTIOA 300
 QY 301 TDMGDGSGTTTAVAVVEILDANDNAPMFDQPKYEAHVPENAVGVHVEVORLVTDLDPNSP 360
 DB 301 TDMGDGSGTTTAVAVVEILDANDNAPMFDQPKYEAHVPENAVGVHVEVORLVTDLDPNSP 360
 QY 361 AWRATYILMGDDGDGHFTITTHPESNQILITTRKGLDFAKNQHTLYVEVTNPAFVLKL 420
 DB 361 AWRATYILMGDDGDGHFTITTHPESNQILITTRKGLDFAKNQHTLYVEVTNPAFVLKL 420
 QY 421 PTSTATIVHVEDVNEAPVFPBPSKVVEVQSGIPTGEPVCVYTAEDPDKENQKISYRIILR 480
 DB 421 PTSTATIVHVEDVNEAPVFPBPSKVVEVQSGIPTGEPVCVYTAEDPDKENQKISYRIILR 480
 QY 481 DPAGWLAMDPSGQVTAAGTGLDREDEQFVRNNIYEVNVLAMONGSPPTGTGTLTLTLLD 540
 DB 481 DPAGWLAMDPSGQVTAAGTGLDREDEQFVRNNIYEVNVLAMONGSPPTGTGTLTLTLLD 540
 QY 541 VNHGVPPEPRQITICNQSPVRHLNITDKLSPHTSPFOAQLTDDSDIYWTAEVNEEGD 600
 DB 541 VNHGVPPEPRQITICNQSPVRHLNITDKLSPHTSPFOAQLTDDSDIYWTAEVNEEGD 600
 QY 601 TVVLSLKKFLKQDITYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFLLPVL 660
 DB 601 TVVLSLKKFLKQDITYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFLLPVL 660
 QY 661 GAVLALLFLLLVLLVLLVRKKRKIKEPILLPEDDTRDNVFFYBEGGGEEDQDYDITQLHR 720
 DB 661 GAVLALLFLLLVLLVLLVRKKRKIKEPILLPEDDTRDNVFFYBEGGGEEDQDYDITQLHR 720
 QY 721 GLEAPBEVLRNDVAPTIIPTPMYRPRPANDPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
 DB 721 GLEAPBEVLRNDVAPTIIPTPMYRPRPANDPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
 QY 781 DYEGSGSDAASLSLTSSASDQDQDYDYLNEWGSRFKKLADMYGGGEDD 829
 DB 781 DYEGSGSDAASLSLTSSASDQDQDYDYLNEWGSRFKKLADMYGGGEDD 829
 RESULT 5
 AAM50864

ID XX AAM50864 standard; protein; 829 AA.
 AC XX AAM50864;
 DT XX 07-MAY-2002 (first entry)
 DE XX Cadherin 3, basal cell marker.
 KW XX Cadherin 3; P-cadherin; human; basal cell; marker; breast cancer; tumour;
 KW XX diagnosis; prognosis; classification; gene therapy.
 OS XX Homo sapiens.
 XX WO200208765-A2.
 PN 31-JAN-2002.
 PD 26-JUL-2001; 2001WO-US023843.
 XX 26-JUL-2000; 2000US-0220967P.
 PR (STRD) UNIV STANFORD.
 PA (GENO-) APPLIED GENOMICS INC.
 XX Botstein D, Brown PO, Perou C, Ross D, Van De Rijn M, Ring B;
 PI Seitz R;
 XX WPI; 2002-206094/26.
 DR Classifying and treating breast tumors associated with the expression and
 PT activity of cadherin 3 or P-cadherin, matrix metalloproteinase 14 and/or
 PT cadherin EGF LAG seven-pass G-type receptor 2 or EGF-like Domain,
 PT Multiple 2.
 XX Claim 1; Fig 1A; 741pp; English.
 CC The present sequence is that of human cadherin 3 (or P-cadherin). The
 CC invention provides new reagents and methods for the management (e.g.
 CC detection, classification, provision of diagnostic and prognostic
 CC information, treatment, etc.) of breast cancer. cDNA microarray
 CC technology was used to identify genes whose expression profile across a
 CC large group of tumour samples correlated with that of cytokeratin 5 and
 CC cytokeratin 17, markers for basal cells of the normal mammary lactation
 CC gland. Tumours that express cytokeratin 5/6 and/or 17 have a poor
 CC prognosis relative to tumours overall. The basal marker genes identified
 CC included those encoding cadherin 3, matrix metalloproteinase 14 (see
 CC AAM50866), and cadherin EGF LAG seven-pass G-type receptor 2 (see
 CC AAM50866). Detection of these markers allows classification of a tumour
 CC sample as belonging to a basal tumour subclass. By providing reagents
 CC that may reliably be used to classify tumours as belonging to a basal
 CC subclass, the invention enables a variety of methods for improving
 CC therapeutic options for patients with breast cancer and for
 CC individualising therapy. Potential therapeutic agents include compounds
 CC that modulate breast basal cell marker genes or that modulate
 CC polypeptides encoded by these genes. In particular, therapeutic agents
 CC include antibodies that specifically bind to the basal marker proteins
 CC and which recognise basal cells of normal mammary lactation glands. The
 CC antibodies can be used either by themselves or when conjugated to, or
 CC delivered with, another molecule such as a toxic compound. The likelihood
 CC that cadherin 3 is membrane-bound makes it an attractive candidate for
 CC antibody therapeutics
 XX
 SQ Sequence 829 AA;
 Query Match 100.0%; Score 829; DB 5; Length 829;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MGLPRGLASLLLLQVCWLOCAASECRAVFAEVTLEAGGAEQPGQALGKVFVCGPG 60
 Db 1 MGLPRGLASLLLLQVCWLOCAASECRAVFAEVTLEAGGAEQPGQALGKVFVCGPG 60
 QY 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKRILLRHRKRDWVAPISVPENG 120

Db 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKRILLRHRKRDWVAPISVPENG 120
 QY 121 KGFPQRLNQLKSNKORDTKIFYSITGPGADSPPEGVFAVEKETGWLKLLKPLDREIAK 180
 Db 121 KGFPQRLNQLKSNKORDTKIFYSITGPGADSPPEGVFAVEKETGWLKLLKPLDREIAK 180
 QY 181 YELFGHVAENGASVEDPMNISIIIVTDNDHDKPKFTQDTFRGSVLGVLPGTSVMQVAT 240
 Db 181 YELFGHVAENGASVEDPMNISIIIVTDNDHDKPKFTQDTFRGSVLGVLPGTSVMQVAT 240
 QY 241 DEDDAIYTVNGVAVSIHSGQPKDHLMTFTHRSSTISVSSGLDREKVPYITLTIQA 300
 Db 241 DEDDAIYTVNGVAVSIHSGQPKDHLMTFTHRSSTISVSSGLDREKVPYITLTIQA 300
 QY 301 TDMGDSSTTAVAVVEILDANDNAPFDPQKYEAVPENAVGHEVORLVTDLDPNSP 360
 Db 301 TDMGDSSTTAVAVVEILDANDNAPFDPQKYEAVPENAVGHEVORLVTDLDPNSP 360
 QY 361 AWRATYLINGGDDGDHFTITTHPESNQGLTTRKGLDFAKNQHTLYVEVTNEAPFVKL 420
 Db 361 AWRATYLINGGDDGDHFTITTHPESNQGLTTRKGLDFAKNQHTLYVEVTNEAPFVKL 420
 QY 421 PTSTATIVHVEDVNEAPVFPSPKVEVEGIGTGPVVCVTAEDPDKENOKISYRILR 480
 Db 421 PTSTATIVHVEDVNEAPVFPSPKVEVEGIGTGPVVCVTAEDPDKENOKISYRILR 480
 QY 481 DPAGWLAMDPSGOVTAVGLTDRDEQFVRNNIYEVWVLANDNGSPPTTGTGLLLTLID 540
 Db 481 DPAGWLAMDPSGOVTAVGLTDRDEQFVRNNIYEVWVLANDNGSPPTTGTGLLLTLID 540
 QY 541 VNDHGVPPEPRQITICNQSPVRVHLNITDKLSPHTSPPQALTTDSDIYWTAEVNEEGD 600
 Db 541 VNDHGVPPEPRQITICNQSPVRVHLNITDKLSPHTSPPQALTTDSDIYWTAEVNEEGD 600
 QY 601 TVVLSLKKFKQDQDYDVHLSLSDHGNKEQLTVIRATVCDCHGVETCPGPKWKGFIPLVL 660
 Db 601 TVVLSLKKFKQDQDYDVHLSLSDHGNKEQLTVIRATVCDCHGVETCPGPKWKGFIPLVL 660
 QY 661 GAVLALLFLLLVLLVLLVLRKKRKEIPELLELLELLELLELLELLELLELLELLELLEL 720
 Db 661 GAVLALLFLLLVLLVLLVLRKKRKEIPELLELLELLELLELLELLELLELLELLELLEL 720
 QY 721 GLEAPPEVWLVNDVAPITIIPTMYRPPANPDEIGNFIENLKAANTDPTAPPYDILLVF 780
 Db 721 GLEAPPEVWLVNDVAPITIIPTMYRPPANPDEIGNFIENLKAANTDPTAPPYDILLVF 780
 QY 781 DYEGSGDAASLSLTSASDQDQDYDLNWSGRFKKLADMYGGGEDD 829
 Db 781 DYEGSGDAASLSLTSASDQDQDYDLNWSGRFKKLADMYGGGEDD 829
 RESULT 6
 ABP58357
 ID ABP58357 standard; protein; 829 AA.
 AC ABP58357;
 XX 07-APR-2003 (first entry)
 DT Human P-cadherin.
 XX P-cadherin; human; colon cancer; colorectal cancer; cytostatic;
 KW gene therapy.
 OS Homo sapiens.
 XX WO200297395-A2.
 PN 05-DEC-2002.
 PD 31-MAY-2002; 2002WO-US017109.
 PF
 XX

PR 31-MAY-2001; 2001US-0294225P.
 XX (CHIR) CHIRON CORP.
 PA Reinhard C, Klinger J, Jefferson AB, Escobedo J, Randazo F;
 PI Winter J, Goodson R;
 XX N-PSDB; ABZ24736.
 XX WPI; 2003-140501/13.
 DR N-PSDB; ABZ24736.
 XX
 XX Inhibiting migration and proliferation of P-cadherin expressing cancer
 PT for treating cancer, especially digestive cancer, characterized by
 PT overexpression of P-cadherin, involves administering a P-cadherin
 PT antagonist.
 XX
 PS Disclosure; Page 10-11; 129pp; English.
 XX
 CC The present sequence is the protein sequence for human P-cadherin
 CC (placental cadherin), a calcium-dependent cellular adhesion protein. The
 CC invention provides methods of treating or diagnosing cancers involving P-
 CC cadherin expression using ligands that target P-cadherin, especially
 CC human anti-P-cadherin antibodies. A claimed method of treating a cancer
 CC characterised by the overexpression and/or upregulation of P-cadherin
 CC comprises the administration of a P-cadherin antagonist, optionally
 CC conjugated to a therapeutic agent. The migration, adhesion and/or
 CC proliferation of the cancer is inhibited, and the method is especially
 CC useful for treating or preventing a digestive cancer such as colon or
 CC colorectal cancer. The antagonists may be an anti-P-cadherin antibody or
 CC its fragment, a ribozyme or antisense oligonucleotide. A transgenic
 CC animal that expresses a recombinant (human) antibody that specifically
 CC binds P-cadherin is also claimed. A method of determining the presence of
 CC cancer involves determining the expression level of P-cadherin in a cell
 CC sample. Also provided are screens for identifying anti-P-cadherin
 CC antibodies have therapeutic activity
 XX
 XX Sequence 829 AA;
 Query Match 100.0%; Score 829; DB 6; Length 829;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MGLPRGLASLLLLQVCLQCAASEPCRAVFREAEVTLLEAGGAOEPPGALGVFMGCPG 60
 DB 1 MGLPRGLASLLLLQVCLQCAASEPCRAVFREAEVTLLEAGGAOEPPGALGVFMGCPG 60
 QY 61 QEPALFSTNDFTVRNGETVQERRSLKERNPLKIPPSKRIILRRKRDWVAPIVSPENG 120
 DB 61 QEPALFSTNDFTVRNGETVQERRSLKERNPLKIPPSKRIILRRKRDWVAPIVSPENG 120
 QY 121 KGFPORLNQKSNKDRDTKIFYSITPGADSPPEGFAVEKETGWLINLKPDLREEIAK 180
 DB 121 KGFPORLNQKSNKDRDTKIFYSITPGADSPPEGFAVEKETGWLINLKPDLREEIAK 180
 QY 181 YELFCHAVSNGASVEDPMNISIIVTDQNDHKPKTQDTFRGSVLEGLVPGTSVMQVAT 240
 DB 181 YELFCHAVSNGASVEDPMNISIIVTDQNDHKPKTQDTFRGSVLEGLVPGTSVMQVAT 240
 QY 241 DEDDALTYTNGVAVYSIHQSEPKDPDLMTFTIHRSTGTISVSSGLDREKVPVYLTITQA 300
 DB 241 DEDDALTYTNGVAVYSIHQSEPKDPDLMTFTIHRSTGTISVSSGLDREKVPVYLTITQA 300
 QY 301 TMDMGDSSTTAVAVVEILLDANDNAPMDPPQYEAHVPENAVGVHVEQRLTVDLDAPNSP 360
 DB 301 TMDMGDSSTTAVAVVEILLDANDNAPMDPPQYEAHVPENAVGVHVEQRLTVDLDAPNSP 360
 QY 361 ANRATYLMGDDGDGHFTTITHPESNOGLITTRKGLDPEAKNQHTLYVEVNEAPFVLKL 420
 DB 361 ANRATYLMGDDGDGHFTTITHPESNOGLITTRKGLDPEAKNQHTLYVEVNEAPFVLKL 420
 QY 421 PTSTATIVVHVEDVNEAPVFPVPPSKVVEQIGPTGEPVCVYTABDPKQNKISYRILR 480
 DB 421 PTSTATIVVHVEDVNEAPVFPVPPSKVVEQIGPTGEPVCVYTABDPKQNKISYRILR 480

QY 481 DPAGWLAMPDPSGGQVTAAGTLDREDEQFVRNNIYEVWVLAMDNQSPPTTGTGTLTLLID 540
 DB 481 DPAGWLAMPDPSGGQVTAAGTLDREDEQFVRNNIYEVWVLAMDNQSPPTTGTGTLTLLID 540
 QY 541 VNDHGVPPEPRQIITICNQSPVRHVNLITDKDLSHTSPFOAQLTDDSDIYWTAEVNEGD 600
 DB 541 VNDHGVPPEPRQIITICNQSPVRHVNLITDKDLSHTSPFOAQLTDDSDIYWTAEVNEGD 600
 QY 601 TWLSLKKFLKQDQYDVHLSLSDHGNKEQLTVIRATVCDCHGVETCPGPKWGGFILPVL 660
 DB 601 TWLSLKKFLKQDQYDVHLSLSDHGNKEQLTVIRATVCDCHGVETCPGPKWGGFILPVL 660
 QY 661 GAVIALFLLLVLLVLLVLRKKIKPEPLLPEDDTRDNVYFYGEGGEEPDQYDITQLHR 720
 DB 661 GAVIALFLLLVLLVLLVLRKKIKPEPLLPEDDTRDNVYFYGEGGEEPDQYDITQLHR 720
 QY 721 GLEARPEVLENDVAPITIIPTMYRPRPANDPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
 DB 721 GLEARPEVLENDVAPITIIPTMYRPRPANDPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
 QY 781 DYEGSGSDAASLSLTSASDQDQDYDLNWSGRFPKLADMYGGGDD 829
 DB 781 DYEGSGSDAASLSLTSASDQDQDYDLNWSGRFPKLADMYGGGDD 829
 RESULT 7
 ADC15497
 ID ADC15497 standard; protein; 829 AA.
 XX
 AC ADC15497;
 XX
 DT 18-DEC-2003 (first entry)
 XX
 DE Human basal cell marker Cadherin-3.
 XX
 KW Human; basal cell marker; cadherin-3; tumour; P-cadherin;
 KW matrix metalloproteinase 14;
 KW cadherin EGF LAG seven-pass G-type receptor 2;
 KW EGF-like domain Multiple 2; mammary lactation gland; breast tumour;
 KW cancer; cytostatic.
 XX
 OS Homo sapiens.
 XX
 PN US2003086934-A1.
 XX
 PD 08-MAY-2003.
 XX
 PF 26-JUL-2001; 2001US-0091849.
 XX
 PR 26-JUL-2000; 2000US-0220967P.
 XX
 PA (BOTS/) BOTSTEIN D.
 PA (BROW/) BROWN P O.
 PA (PERO/) PEROU C M.
 PA (RING/) RING B.
 PA (ROSS/) ROSS D.
 PA (SEIT/) SEITZ R.
 PA (VRIJ/) VAN DE RIJN J M.
 XX
 PI Borstein D, Brown PO, Perou CM, Ring B, Ross D, Seitz R;
 PI Van De Rijn JM;
 XX
 DR WPI; 2003-786931/74.
 XX
 PT Classifying breast tumor by detecting expression or activity of gene
 PT encoding cadherin 3, matrix metalloproteinase 14, cadherin EGF LAG seven-
 PT pass G-type receptor 2 in tumor sample.
 XX
 PS Claim 1; SEQ ID NO 1; 89pp; English.
 XX
 CC The invention relates to classifying (M1) a tumour comprising detecting
 CC expression or activity of a gene encoding cadherin 3 or P-cadherin
 CC ,matrix metalloproteinase 14 , or cadherin EGF LAG seven-pass G-type

CC receptor 2 or EGF-like domain, Multiple 2 (P1-3) or at least two genes
CC chosen from gene encoding (P1), (P2) or (P3) in the tumour sample, and
CC classifying tumour as belonging to a tumour subclass based on results of
CC detecting step. Also included are testing a subject for expression of P1-
CC P3 and providing diagnostic, prognostic, or predictive information based
CC on the detecting step, stratifying the subject for a clinical trial based
CC on the detecting step, or selecting a treatment based on the detecting
CC step, an antibody that specifically binds to an epitope found in (P1),
CC (P2) or (P3) (where the antibody recognises basal cells in normal mammary
CC lactation glands), methods of testing a compound or a combination of
CC compounds for activity against tumours. The method (M1) is useful for
CC classifying a tumour (breast tumour), as belonging to a tumour subclass
CC such as basal tumour subclass. The antibody is useful for treating a
CC subject in need of treatment for cancer. The tumour is a breast tumour,
CC and where the method further involves identifying the tumour as belonging
CC to a basal tumour subclass. The antibody is conjugated with a toxic
CC molecule. The present sequence is human Cadherin-3.
XX
SQ Sequence 829 AA;
Query Match 100.0%; Score 829; DB 7; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGLPRGLASLLLLQVCWLCQAASEPCRAVFREAEVTL EAGGAEPGQALGKVFMGCPG 60
DB 1 MGLPRGLASLLLLQVCWLCQAASEPCRAVFREAEVTL EAGGAEPGQALGKVFMGCPG 60
QY 61 QEPALFSTNDNDFTVNGETVQERRSLKERNP LKIPPSKRILRRHKRDWVAPI SPVENG 120
DB 61 QEPALFSTNDNDFTVNGETVQERRSLKERNP LKIPPSKRILRRHKRDWVAPI SPVENG 120
QY 121 KGPFPQRLNQLKSKNDRTKIFYSITGPGADSPPEGVFAVEKETGMLL LKPLDREETAK 180
DB 121 KGPFPQRLNQLKSKNDRTKIFYSITGPGADSPPEGVFAVEKETGMLL LKPLDREETAK 180
QY 181 YELFGHAVENGASVEDPMNISIIVTDQNDHKPKFTQDTRFGSLVGLPGT SVMQVAT 240
DB 181 YELFGHAVENGASVEDPMNISIIVTDQNDHKPKFTQDTRFGSLVGLPGT SVMQVAT 240
QY 241 DEDDAITYNGVAVYHSGQEPKDPDLMTFTHRSSTGTSISVSSGLDREKVP EYTLTQA 300
DB 241 DEDDAITYNGVAVYHSGQEPKDPDLMTFTHRSSTGTSISVSSGLDREKVP EYTLTQA 300
QY 301 TMDGCGSTTTAVAVVEILDANDNAMPDPOKYEAVHPENAVGHEVORLT VTDLPANSP 360
DB 301 TMDGCGSTTTAVAVVEILDANDNAMPDPOKYEAVHPENAVGHEVORLT VTDLPANSP 360
QY 361 AWRATYLMGGDDGDFHTTTHPESNQGLTTRKGLDFAKQHTLYVEVTNEAPV LXL 420
DB 361 AWRATYLMGGDDGDFHTTTHPESNQGLTTRKGLDFAKQHTLYVEVTNEAPV LXL 420
QY 421 PTSTATIVHVEDVNEAPVFPVPSKVVEQGIPTGEPVCVYTAEPD KENQISYRILR 480
DB 421 PTSTATIVHVEDVNEAPVFPVPSKVVEQGIPTGEPVCVYTAEPD KENQISYRILR 480
QY 481 DPAGWLAMPDQGVTAAGTLDREDOFVRNNIYEVWVLAMDNQSPPTGTG TLLTLTD 540
DB 481 DPAGWLAMPDQGVTAAGTLDREDOFVRNNIYEVWVLAMDNQSPPTGTG TLLTLTD 540
QY 541 VNDHGPVPEPRQITCNQSPVRHLNITDKLSPHTSPFOAQLTDSDIYWTAEVNEEGD 600
DB 541 VNDHGPVPEPRQITCNQSPVRHLNITDKLSPHTSPFOAQLTDSDIYWTAEVNEEGD 600
QY 601 TVVLSLKFFLKQDTYVHLSLSDHGNKEQLTVIRATVCDCHGHEVCPGFWKGGF LPLVL 660
DB 601 TVVLSLKFFLKQDTYVHLSLSDHGNKEQLTVIRATVCDCHGHEVCPGFWKGGF LPLVL 660
QY 661 GAVLALLFLLLVLLLVLLVRRKKRKEKILLPDLDDTRDNVNFYFEGGGEEDQD YITQLHR 720
DB 661 GAVLALLFLLLVLLLVLLVRRKKRKEKILLPDLDDTRDNVNFYFEGGGEEDQD YITQLHR 720
QY 721 GLEARPEVLRNDVAPTIITPTMYRPRPANPDEIGNFIIENLKAANTDPTAPPYDTLLVF 780

DB 721 GLEARPEVLRNDVAPTIITPTMYRPRPANPDEIGNFIIENLKAANTDPTAPPYDTLLVF 780
QY 781 DYEGSGDAASLSLTSSASDQDQDYDLNMGSRFKKLADMYGGGDD 829
DB 781 DYEGSGDAASLSLTSSASDQDQDYDLNMGSRFKKLADMYGGGDD 829
RESULT 8
ADD14190
ID ADD14190 standard; protein; 829 AA.
XX
AC ADD14190;
XX
DT 01-JAN-2004 (first entry)
XX
DE Human src biomarker polypeptide SEQ ID NO:379.
XX
KW predictor set; protein tyrosine kinase activity modulator;
KW protein tyrosine kinase pathway; protein tyrosine kinase; cytostatic;
KW gene therapy; drug sensitivity; genetic profile; cancer; human.
XX
OS Homo sapiens.
XX
PN WO2003062395-A2.
XX
PD 31-JUL-2003.
XX
PF 17-JAN-2003; 2003WO-US001981.
XX
PR 18-JAN-2002; 2002US-0350061P.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI Huang P, Fairchild CR, Lee FY, Shaw P;
XX
DR WPI; 2003-636735/60.
DR N-PSDB; ADD14795.
XX
PT New polynucleotides and polypeptides for predicting the activity of
PT compounds that interact with protein tyrosine kinases and/or protein
PT tyrosine kinase pathways.
XX
PS Claim 10; SEQ ID NO 379; 139pp; English.
XX
CC The present invention describes a predictor set comprising a plurality of
CC polynucleotides or polypeptides whose expression pattern is predictive of
CC the response of cells to treatment with a compound that modulates protein
CC tyrosine kinase activity or members of the protein tyrosine kinase
CC pathway. Also described: (1) predicting whether a compound is capable of
CC modulating the activity of cells, comprising obtaining a sample of cells,
CC determining whether the cells express a plurality of markers, and
CC correlating the expression of the markers to the compound's ability to
CC modulate the activity of the cells; (2) a plurality of cell lines for
CC identifying polynucleotides and polypeptides whose expression levels
CC correlate with compound sensitivity or resistance of cells associated
CC with a disease state; and (3) identifying polynucleotides and
CC polypeptides that predict compound sensitivity or resistance of cells
CC associated with a disease state, comprising subjecting the plurality of
CC cell lines to one or more compounds, analysing the expression pattern of
CC a microarray of polynucleotides or polypeptides, and selecting
CC polynucleotides or polypeptides that predict the sensitivity or
CC resistance of cells associated with a disease state by using the
CC expression pattern of the microarray. The polynucleotides and
CC polypeptides have cytostatic activities, and can be used in gene therapy.
CC The polynucleotides and polypeptides are useful in predicting the
CC activity of compounds that interact with protein tyrosine kinases and/or
CC protein tyrosine kinase pathways. These may be used in determining drug
CC sensitivity in patients to allow the development of individualized
CC genetic profiles which aid in treating diseases and disorders (e.g.
CC cancer) based on patient response at a molecular level. The present
CC sequence is used in the exemplification of the present invention.
XX

SQ Sequence 829 AA;
 Query Match 100.0%; Score 829; DB 7; Length 829;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPRGLASILLQVCWLCQAASEPCRAVFAEAVTLEAGGAEQEPGQALGVFMGCPG 60
 DB 1 MGLPRGLASILLQVCWLCQAASEPCRAVFAEAVTLEAGGAEQEPGQALGVFMGCPG 60
 QY 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDWVAPISVPENG 120
 DB 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDWVAPISVPENG 120
 QY 121 KGPPFQRLNQLKSNKORDTKIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDREBIAK 180
 DB 121 KGPPFQRLNQLKSNKORDTKIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDREBIAK 180
 QY 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSVLEGLPGTSVMQVAT 240
 DB 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSVLEGLPGTSVMQVAT 240
 QY 241 DEDDAIYTVNGVAVSHSQEPKDPHDLMTIHRSTGTISVISGLDREKVPYTLTIOA 300
 DB 241 DEDDAIYTVNGVAVSHSQEPKDPHDLMTIHRSTGTISVISGLDREKVPYTLTIOA 300
 QY 301 TMDGSGSTTTAVAVVEILDANDNAPMFDQPKYEAHVPENAVGHVEQRLTVDLDAPNSP 360
 DB 301 TMDGSGSTTTAVAVVEILDANDNAPMFDQPKYEAHVPENAVGHVEQRLTVDLDAPNSP 360
 QY 361 AWRATYILMGDDGDHPTIITHPESNQILTTTRKGLDFAKNQHTLYVEVNEAPVVLK 420
 DB 361 AWRATYILMGDDGDHPTIITHPESNQILTTTRKGLDFAKNQHTLYVEVNEAPVVLK 420
 QY 421 PTSTATIIVHVEDVNEAPVFPVSKVEVOEGIPGTCPEVCVYTAEDPDKENOKISVRIIL 480
 DB 421 PTSTATIIVHVEDVNEAPVFPVSKVEVOEGIPGTCPEVCVYTAEDPDKENOKISVRIIL 480
 QY 481 DPAGWLAMDPSQGVTAAGTLDREDSQFVRNNIYEVWVLAMDNQSPPTTGTGLTLTLD 540
 DB 481 DPAGWLAMDPSQGVTAAGTLDREDSQFVRNNIYEVWVLAMDNQSPPTTGTGLTLTLD 540
 QY 541 VNDHGVPEPQITICNOSPVRVNIATDKOLSPHTSPFQAQLTDDSDIYWTAEVNEED 600
 DB 541 VNDHGVPEPQITICNOSPVRVNIATDKOLSPHTSPFQAQLTDDSDIYWTAEVNEED 600
 QY 601 TVVLSLKKFLKQDYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKGGFILPVL 660
 DB 601 TVVLSLKKFLKQDYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKGGFILPVL 660
 QY 661 GAVLALLFLLVLLVLRKKRKIKPELLIPEDDTRDNVFIYEGEGGEDQDYDTQLHR 720
 DB 661 GAVLALLFLLVLLVLRKKRKIKPELLIPEDDTRDNVFIYEGEGGEDQDYDTQLHR 720
 QY 721 GLEAREVILRNDVAPTIPTMYRPRRPNDEIGNFIENKAANTDPTAPPYDTLLVF 780
 DB 721 GLEAREVILRNDVAPTIPTMYRPRRPNDEIGNFIENKAANTDPTAPPYDTLLVF 780
 QY 781 DYEGSGDAASLSLTSASDQDDQDYLYLNEWGSRFKKLADMYGGGEDD 829
 DB 781 DYEGSGDAASLSLTSASDQDDQDYLYLNEWGSRFKKLADMYGGGEDD 829

RESULT 9
 ABF54683
 ID ABF54683 standard; protein; 829 AA.
 XX
 AC ABF54683;
 XX
 DT 30-DEC-2002 (first entry)
 XX
 DE Metastatic colorectal cancer-associated polypeptide.
 XX

KW Colorectal cancer; metastasis; differential expression; cytostatic;
 XX diagnosis; gene therapy; vaccine.
 OS Homo sapiens.
 XX
 PN WO200268677-A2.
 XX
 PD 06-SEP-2002.
 XX
 PF 27-FEB-2002; 2002WO-US006001.
 XX
 PR 27-FEB-2001; 2001US-0272206P.
 PR 02-APR-2001; 2001US-0281149P.
 PR 17-APR-2001; 2001US-0284555P.
 XX
 XX (EOSB-) EOS BIOTECHNOLOGY INC.
 PA (UYCA-) UNIV CASE WESTERN RESERVE.
 XX
 XX Mack DH, Markowitz SD;
 PI
 XX WPI; 2002-698677/75.
 DR N-PSDB; ABQ81547.
 XX
 PT New genes that are up- or down-regulated in colorectal cancer, useful for
 PT diagnosing colorectal cancer in a subject, or for identifying modulators
 PT of colorectal cancer-associated proteins and genes for treating
 PT colorectal cancer.
 PT
 XX
 PS Claim 8; Page 245; 260pp; English.
 XX
 CC The present sequence is the protein sequence of a human polypeptide
 CC encoded by a gene that exhibits increased expression in colon cancer-
 CC derived metastases compared to normal colon tissue. It is an example of
 CC claimed polypeptides that are encoded by genes which are differentially
 CC expressed in metastatic colorectal cancer cells. Such polypeptides are
 CC useful in diagnostic and prognostic assays, for raising antibodies useful
 CC e.g. in immunotherapy, and in screening for modulator compounds of
 CC therapeutic value
 XX
 SQ Sequence 829 AA;

Query Match 87.8%; Score 728; DB 5; Length 829;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 828; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGLPRGLASILLQVCWLCQAASEPCRAVFAEAVTLEAGGAEQEPGQALGVFMGCPG 60
 DB 1 MGLPRGLASILLQVCWLCQAASEPCRAVFAEAVTLEAGGAEQEPGQALGVFMGCPG 60
 QY 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDWVAPISVPENG 120
 DB 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDWVAPISVPENG 120
 QY 121 KGPPFQRLNQLKSNKORDTKIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDREBIAK 180
 DB 121 KGPPFQRLNQLKSNKORDTKIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDREBIAK 180
 QY 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSVLEGLPGTSVMQVAT 240
 DB 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSVLEGLPGTSVMQVAT 240
 QY 241 DEDDAIYTVNGVAVSHSQEPKDPHDLMTIHRSTGTISVISGLDREKVPYTLTIOA 300
 DB 241 DEDDAIYTVNGVAVSHSQEPKDPHDLMTIHRSTGTISVISGLDREKVPYTLTIOA 300
 QY 301 TMDGSGSTTTAVAVVEILDANDNAPMFDQPKYEAHVPENAVGHVEQRLTVDLDAPNSP 360
 DB 301 TMDGSGSTTTAVAVVEILDANDNAPMFDQPKYEAHVPENAVGHVEQRLTVDLDAPNSP 360
 QY 361 AWRATYILMGDDGDHPTIITHPESNQILTTTRKGLDFAKNQHTLYVEVNEAPVVLK 420
 DB 361 AWRATYILMGDDGDHPTIITHPESNQILTTTRKGLDFAKNQHTLYVEVNEAPVVLK 420

QY 421 PSTATIVVHVEDVNEAPVFPVPPSKVVEVQEGIPGTGEPVVCVVTABDPKQKISYRILR 480
 DB 421 PSTATIVVHVEDVNEAPVFPVPPSKVVEVQEGIPGTGEPVVCVVTABDPKQKISYRILR 480
 QY 481 DPAGWLAMPDPSGQVAVGTGLDREDEQFVRNNIYEVWVLAMDNQSPPTTGTGTLTLLID 540
 DB 481 DPAGWLAMPDPSGQVAVGTGLDREDEQFVRNNIYEVWVLAMDNQSPPTTGTGTLTLLID 540
 QY 541 VNDHGVPPEPRQITICNQSPVRHVLNITDKDLSPTSPFQAQLTDDSDIYVTAENVNEGD 600
 DB 541 VNDHGVPPEPRQITICNQSPVRHVLNITDKDLSPTSPFQAQLTDDSDIYVTAENVNEGD 600
 QY 601 TVVLSLKXFLKQDITYDVHLSLSDHGNKEQLTVIRATVCDCHGVETCFGPNKGGFIPVL 660
 DB 601 TVVLSLKXFLKQDITYDVHLSLSDHGNKEQLTVIRATVCDCHGVETCFGPNKGGFIPVL 660
 QY 661 GAVLALLFLLVLLVLLVLRKKIKFLLLPEDDTRDNVFFYGGEGGEDQDIDITQLHR 720
 DB 661 GAVLALLFLLVLLVLLVLRKKIKFLLLPEDDTRDNVFFYGGEGGEDQDIDITQLHR 720
 QY 721 GLEAPPEVLRNDVAPTPTPTMYPRPANPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
 DB 721 GLEAPPEVLRNDVAPTPTPTMYPRPANPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
 QY 781 DYEGSGDAASLSSTSSASDQDQDYDYLNWGSRFKKLADMYGGGEDD 829
 DB 781 DYEGSGDAASLSSTSSASDQDQDYDYLNWGSRFKKLADMYGGGEDD 829

RESULT 10
 ID ABR58670 standard; protein; 829 AA.
 AC ABR58670;
 DT 09-JUL-2003 (first entry)
 XX Human cancer related protein SEQ ID NO:327.
 DE Human; cancer; diagnosis; screening; modulator; leukaemia; ischaemia;
 KW heart disease; atherosclerosis; endometriosis.
 XX Homo sapiens.
 XX WO2003025138-A2.
 PN 27-MAR-2003.
 PD 17-SEP-2002; 2002WO-US029560.
 PF 17-SEP-2001; 2001US-0323469P.
 PR 20-SEP-2001; 2001US-0323887P.
 PR 13-NOV-2001; 2001US-0350666P.
 PR 08-FEB-2002; 2002US-0355145P.
 PR 08-FEB-2002; 2002US-0355257P.
 PR 12-APR-2002; 2002US-0372246P.
 XX (E0SB-) EOS BIOTECHNOLOGY INC.
 PA Afar D, Aziz N, Gish KC, Hevezi PA, Mack DH, Wilson KE;
 XX Zlotnik A;
 PI WPI: 2003-354600/33.
 XX N-PSDB; ACC72821.

XX New genes that are up-regulated or down-regulated in cancers, useful as
 PT markers for diagnosing e.g. cancer, ischaemia or heart diseases, or as
 PT therapeutic targets for screening drugs for treating these diseases.

XX Claim 12; Page 757; 767pp; English.

XX The present invention describes an isolated nucleic acid molecule, which
 CC comprises the sequence of any of the genes that are up-regulated or down-

CC regulated in specific cancers (e.g. about 1031 genes up-regulated in
 CC acute lymphocytic leukemia). ACC72841 to ACC72860 represent cancer
 CC related gene nucleotide sequences which encode the proteins given in
 CC ABR58521 to ABR58709. Also described: (1) determining the presence or
 CC absence of a pathological cell in a patient; (2) an expression vector
 CC comprising a nucleic acid molecule described above; (3) a host cell
 CC comprising the vector; (4) an isolated polypeptide, which is encoded by
 CC the nucleic acid; (5) an antibody that specifically binds the polypeptide
 CC of (4); (6) specifically targeting a compound to a pathological cell in a
 CC patient by administering to the patient the antibody above; and (7) a
 CC drug screening assay. The nucleic acid is useful as diagnostic markers or
 CC therapeutic targets. In particular, the nucleic acid is useful for
 CC diagnosing a pathology, e.g. cancer (e.g. cancer of the bone marrow,
 CC bladder, brain, breast, cervix, colon/rectum, kidney, lung, ovary,
 CC pancreas, prostate, skin and uterus), wounds, ischaemia, heart diseases,
 CC atherosclerosis and endometriosis. The nucleic acid is also useful in
 CC drug screening, particularly for identifying agents for treating these
 CC pathologies

XX SQ Sequence 829 AA;

Query Match 87.8%; Score 728; DB 6; Length 829;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 828; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGUPRGPASLLLLQVCWLQCAASEPCRAVFAEAVTLEAGAEQFPQALGVFGCPG 60
 DB 1 MGUPRGPASLLLLQVCWLQCAASEPCRAVFAEAVTLEAGAEQFPQALGVFGCPG 60
 QY 61 QEPALFSTDNDDFTVRNGETVQERRSIKERNPLKIPPSKRILRRHKEDVWVAPISVPENG 120
 DB 61 QEPALFSTDNDDFTVRNGETVQERRSIKERNPLKIPPSKRILRRHKEDVWVAPISVPENG 120
 QY 121 KGPPPRQLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKGTGMLLNKPLDRSEIAK 180
 DB 121 KGPPPRQLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKGTGMLLNKPLDRSEIAK 180
 QY 181 YELFPGHAVSENGASVEDPMNISIIVTQNDHKPKFTQDTPRGSVLPGVSGVMQVAT 240
 DB 181 YELFPGHAVSENGASVEDPMNISIIVTQNDHKPKFTQDTPRGSVLPGVSGVMQVAT 240
 QY 241 DEDDAIYTVNGVAVYSIHSQEPKDPHLMFTIHRSTGTISVISGGLDREKVPYTTTQA 300
 DB 241 DEDDAIYTVNGVAVYSIHSQEPKDPHLMFTIHRSTGTISVISGGLDREKVPYTTTQA 300
 QY 301 TDMGDGSTTTAVAVBEILDANDNAPMFDQKYEAHVPENAVGHEVQRLTVDLAPNSP 360
 DB 301 TDMGDGSTTTAVAVBEILDANDNAPMFDQKYEAHVPENAVGHEVQRLTVDLAPNSP 360
 QY 361 AWRTATYLINGDGDGDHFTITTHPESNOGILTTKGLDPEAKNQHTLYVEVNEAPFVLKL 420
 DB 361 AWRTATYLINGDGDGDHFTITTHPESNOGILTTKGLDPEAKNQHTLYVEVNEAPFVLKL 420
 QY 421 PTSTATIVVHVEDVNEAPVFPVPPSKVVEVQEGIPTGEPVVCVVTABDPKQKISYRILR 480
 DB 421 PTSTATIVVHVEDVNEAPVFPVPPSKVVEVQEGIPTGEPVVCVVTABDPKQKISYRILR 480
 QY 481 DPAGWLAMPDPSGQVAVGTGLDREDEQFVRNNIYEVWVLAMDNQSPPTTGTGTLTLLID 540
 DB 481 DPAGWLAMPDPSGQVAVGTGLDREDEQFVRNNIYEVWVLAMDNQSPPTTGTGTLTLLID 540
 QY 541 VNDHGVPPEPRQITICNQSPVRHVLNITDKDLSPTSPFQAQLTDDSDIYVTAENVNEGD 600
 DB 541 VNDHGVPPEPRQITICNQSPVRHVLNITDKDLSPTSPFQAQLTDDSDIYVTAENVNEGD 600
 QY 601 TVVLSLKXFLKQDITYDVHLSLSDHGNKEQLTVIRATVCDCHGVETCFGPNKGGFIPVL 660
 DB 601 TVVLSLKXFLKQDITYDVHLSLSDHGNKEQLTVIRATVCDCHGVETCFGPNKGGFIPVL 660
 QY 661 GAVLALLFLLVLLVLLVLRKKIKFLLLPEDDTRDNVFFYGGEGGEDQDIDITQLHR 720
 DB 661 GAVLALLFLLVLLVLLVLRKKIKFLLLPEDDTRDNVFFYGGEGGEDQDIDITQLHR 720

QY 721 GLEAPVVLNDVAPTIPTMYRPRPANPDIGNFIENLKAANTDPTAPPYDTLLVF 780
 DB |||||
 QY 781 DYEGSGDAASLSLSSASDQDDYDLNMGSRFKKLADMYGGGDD 829
 DB |||||
 QY 781 DYEGSGDAASLSLSSASDQDDYDLNMGSRFKKLADMYGGGDD 829
 DB |||||

RESULT 11
 ABUS6670
 ID ABUS6670 standard; protein; 829 AA.
 XX AC
 XX AC ABUS6670;
 XX
 DT 02-APR-2003 (first entry)
 DE Lung cancer-associated polypeptide #27.
 XX
 KW Lung cancer-associated polypeptide; cytostatic; emphysema;
 KW antiinflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;
 KW small cell lung cancer; benign lesion; precancerous lesion; bronchitis;
 KW chronic obstructive pulmonary disease; hypersensitivity pneumonitis;
 KW interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.
 XX
 OS Unidentified.
 XX
 PN WO200286443-A2.
 XX
 PD 31-OCT-2002.
 XX
 XX 18-APR-2002; 2002WO-US012476.
 XX
 PR 16-APR-2001; 2001US-0284770P.
 PR 10-MAY-2001; 2001US-0290492P.
 PR 09-NOV-2001; 2001US-0339245P.
 PR 13-NOV-2001; 2001US-0350666P.
 PR 29-NOV-2001; 2001US-0334370P.
 PR 12-APR-2002; 2002US-0372246P.
 XX
 PA (BOSB-) BOS BIOTECHNOLOGY INC.
 XX
 PI Aziz N, Murray R;
 XX
 DR WFI; 2003-093161/08.
 DR N-PSDB; ABX76155.
 XX
 PT Detecting a lung cancer-associated transcript in a cell from a patient
 PT for treating lung cancer, by contacting a biological sample from the
 PT patient with a polynucleotide that exhibits increased or decreased
 PT expression in lung cancer.
 XX
 XX Claim 27; Page 210; 453pp; English.
 XX
 CC The invention relates to a method for detecting a lung cancer-associated
 CC transcript in a cell from a patient, comprising contacting a biological
 CC sample from the patient with a polynucleotide that selectively hybridises
 CC to a sequence that is at least 80 % identical to a gene that exhibits
 CC increased or decreased expression in lung cancer samples. Lung cancer-
 CC associated polynucleotides and polypeptides are used for identifying a
 CC compound that modulates a lung cancer-associated polypeptide, for
 CC inhibiting proliferation of a lung cancer-associated cell to treat lung
 CC cancer in a patient and for treating a mammal having lung cancer by
 CC administering a modulatory compound identified. The methods are useful
 CC for treating lung cancer, such as small cell lung cancer, non-small cell
 CC lung cancer or other benign or precancerous lesions, e.g. atelectasis,
 CC emphysema, bronchitis, chronic obstructive pulmonary disease, fibrosis,
 CC hypersensitivity pneumonitis, interstitial pulmonary fibrosis, asthma and
 CC bronchiectasis. The genes, polynucleotides and polypeptides are useful
 CC for diagnostic purposes and as targets for screening for therapeutic
 CC compounds that modulate lung cancer, such as antibodies. Sequences
 CC ABUS6408-ABUS6745 represent lung cancer-associated polypeptides of the
 CC invention

XX Sequence 829 AA;
 SQ Query Match 87.8%; Score 728; DB 6; Length 829;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 828; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGLPRGFLASLLLOVCWLOCAASEPCRAVFAEAETLEAGGAEQPGQALGVFVGCPG 60
 DB 1 MGLPRGFLASLLLOVCWLOCAASEPCRAVFAEAETLEAGGAEQPGQALGVFVGCPG 60
 QY 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIPFSKILRRHKRDVWVAPISVPENG 120
 DB 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIPFSKILRRHKRDVWVAPISVPENG 120
 QY 121 KGPPQRLNOLKSNKORDTKIFYISITGPGADSPGEGFAVEKETGMLLNKPLDREIIAK 180
 DB 121 KGPPQRLNOLKSNKORDTKIFYISITGPGADSPGEGFAVEKETGMLLNKPLDREIIAK 180
 QY 181 YELFGHAVSNGASVEDPMNISIIITDQNDHKPKFTQDIFRGSVLEGVLPSTVMQVAT 240
 DB 181 YELFGHAVSNGASVEDPMNISIIITDQNDHKPKFTQDIFRGSVLEGVLPSTVMQVAT 240
 QY 241 DEDDAITYNGVWAYSIIHSQEPKDPHDLMTIHRSTGTISVISSGLDREKVPYTTIQA 300
 DB 241 DEDDAITYNGVWAYSIIHSQEPKDPHDLMTIHRSTGTISVISSGLDREKVPYTTIQA 300
 QY 301 TDMGDGSTTTAVAVVEILDANDNAPMFDPOKYEAHVPENAVGHEVQRLTVDLDPNSP 360
 DB 301 TDMGDGSTTTAVAVVEILDANDNAPMFDPOKYEAHVPENAVGHEVQRLTVDLDPNSP 360
 QY 361 AWRATYILMGDDGDHPTIITHPESNOGILITTRKGLDFAKNOHTLYVEVTNAPFVKL 420
 DB 361 AWRATYILMGDDGDHPTIITHPESNOGILITTRKGLDFAKNOHTLYVEVTNAPFVKL 420
 QY 421 PTSTATIIVHVEDVNEAPVFPVPSKVVEOEGPTGEPVCVYTAEDPDKENQKISVRIILR 480
 DB 421 PTSTATIIVHVEDVNEAPVFPVPSKVVEOEGPTGEPVCVYTAEDPDKENQKISVRIILR 480
 QY 481 DPAGWLAMPDPSGOVTAVGTLDRDEDEQFVRNNTYEVNVLAMDNPSPTTGTGTLTLID 540
 DB 481 DPAGWLAMPDPSGOVTAVGTLDRDEDEQFVRNNTYEVNVLAMDNPSPTTGTGTLTLID 540
 QY 541 VNDHGPVPEPQITICNOSPVRHLNITDKLSPTSPFQAQLTDDSDIYWTAEVNEEGD 600
 DB 541 VNDHGPVPEPQITICNOSPVRHLNITDKLSPTSPFQAQLTDDSDIYWTAEVNEEGD 600
 QY 601 TVVLSLKKFLKQDITYVHLSLSDHGKNEQLTVIRATVCDCHGHVETCPGPKWGGFILPVL 660
 DB 601 TVVLSLKKFLKQDITYVHLSLSDHGKNEQLTVIRATVCDCHGHVETCPGPKWGGFILPVL 660
 QY 661 GAVLALLFLLVLLVLLVLRKKRKEPILLPDDTRDNVFFYEGEGGEDDQDITQLHR 720
 DB 661 GAVLALLFLLVLLVLLVLRKKRKEPILLPDDTRDNVFFYEGEGGEDDQDITQLHR 720
 QY 721 GLEAPVVLNDVAPTIPTMYRPRPANPDIGNFIENLKAANTDPTAPPYDTLLVF 780
 DB 721 GLEAPVVLNDVAPTIPTMYRPRPANPDIGNFIENLKAANTDPTAPPYDTLLVF 780
 QY 781 DYEGSGDAASLSLSSASDQDDYDLNMGSRFKKLADMYGGGDD 829
 DB 781 DYEGSGDAASLSLSSASDQDDYDLNMGSRFKKLADMYGGGDD 829

RESULT 12
 ABUS6670
 ID ABUS6670 standard; protein; 829 AA.
 XX AC
 XX AC ABUS6670;
 XX
 DT 02-APR-2003 (first entry)
 DE Lung cancer-associated polypeptide #263.

DR WPI; 1999-263686/22.
 XX Modulating cadherin-mediated cell adhesion useful for treating
 PT neurological disease and cancer.
 XX
 XX Disclosure; Fig 2; 149pp; English.
 XX
 CC The invention relates to methods using cadherin modulating agents,
 CC particularly peptides containing the sequence HAV or antibodies, for
 CC enhancing or inhibiting cadherin-mediated cell adhesion. The methods can
 CC be used to treat diseases or other conditions characterized by
 CC undesirable cell adhesion or to facilitate drug delivery to a specific
 CC tissue or tumour (such as bladder tumours, ovarian tumours or melanomas).
 CC They can also be used to enhance cell adhesion (e.g. supplement or
 CC replace stitches or to facilitate wound healing). They can also be used
 CC for enhancing and/or directing neurite outgrowth. The methods can also be
 CC used for treating demyelinating neurological disease, e.g. multiple
 CC sclerosis. The methods can also be used for e.g. enhancing drug delivery,
 CC treating cancers (such as carcinomas, leukemia or melanomas), inhibiting
 CC angiogenesis, enhancing adhesion of foreign tissue implanted in a mammal,
 CC inducing apoptosis in cadherin-expressing cells, e.g. for treating
 CC diabetes or rheumatoid arthritis, modulating the immune system, for
 CC preventing pregnancy, increasing vasopermeability, treating spinal cord
 CC injuries or inhibiting synaptic stability
 XX
 XX Sequence 108 AA;

Query Match 13.0%; Score 108; DB 2; Length 108;
 Best Local Similarity 100.0%; Pred. No. 3.4e-93;
 Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 108 DWVAPISVPENGKGPFPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 167
 DB 1 DWVAPISVPENGKGPFPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 60
 QY 168 LLNKPLDREIAKYELFGHVSNGASVEDPMNISIIIVTDQNDHKPKF 215
 DB 61 LLNKPLDREIAKYELFGHVSNGASVEDPMNISIIIVTDQNDHKPKF 108

RESULT 14
 AAY17104
 ID AAY17104 standard; protein; 108 AA.
 AC AAY17104;
 DT 20-SEP-1999 (first entry)
 DE Human P-cadherin EC1 domain sequence.
 XX
 XX Cell adhesion modulation; CAM; synaptic stability; cadherin;
 KW cadherin-mediated adhesion; drug delivery; cell adhesion; tumour;
 KW wound healing; neurite outgrowth.
 XX
 XX Homo sapiens.
 OS
 XX WO9933875-A1.
 FN
 XX 08-JUL-1999.
 PD
 XX 23-DEC-1998; 98WO-CA001207.
 PF
 XX 23-DEC-1997; 97US-00996679.
 PR
 XX (UYMC-) UNIV MCGILL.
 PA Blaschuk OW, Gour BJ;
 XX WPI; 1999-430231/36.
 DR
 XX Cyclic peptide cell adhesion modulating agents, useful for modulating
 PT synaptic stability.
 XX

PS Disclosure; Fig 2; 144pp; English.
 XX
 CC The invention provides cyclic peptide cell adhesion modulating (CAM)
 CC agents that comprises a His-Ala-Val recognition sequence. Also provided
 CC is a method for inhibiting synaptic stability in a mammal that comprises
 CC administering to a mammal a therapeutically effective amount of a CAM
 CC agent that inhibits cadherin-mediated adhesion, where the agent comprises
 CC a cyclic peptide having a peptide ring, and where the sequence His-Ala-
 CC Val is present within the peptide ring. The cyclic peptides are cell
 CC adhesion modulating agents that inhibit cadherin-mediated adhesion. They
 CC can be used in a method for inhibiting synaptic stability in mammals. The
 CC agents can be used to treat diseases or other conditions characterized by
 CC undesirable cell adhesion or to facilitate drug delivery to a specific
 CC tissue or tumour. Alternatively the agents may be used to enhance cell
 CC adhesion (e.g. to supplement or replace stitches or to facilitate wound
 CC healing) or to enhance or direct neurite outgrowth
 XX
 XX Sequence 108 AA;

Query Match 13.0%; Score 108; DB 2; Length 108;
 Best Local Similarity 100.0%; Pred. No. 3.4e-93;
 Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 108 DWVAPISVPENGKGPFPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 167
 DB 1 DWVAPISVPENGKGPFPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 60
 QY 168 LLNKPLDREIAKYELFGHVSNGASVEDPMNISIIIVTDQNDHKPKF 215
 DB 61 LLNKPLDREIAKYELFGHVSNGASVEDPMNISIIIVTDQNDHKPKF 108

RESULT 15
 AAY78184
 ID AAY78184 standard; protein; 108 AA.
 AC AAY78184;
 DT 03-MAY-2000 (first entry)
 DE Human P-cadherin EC1 domain SEQ ID NO:37.
 XX
 XX Cadherin; modulation; cell adhesion recognition; CAR; HAV-BF;
 KW HAV binding motif; antiapoptotic; cytostatic; cell adhesion;
 KW neurite outgrowth; Schwann cell migration; tumour; cancer; metastasis;
 KW apoptosis; angiogenesis; immune system; pregnancy; vasopermeability;
 KW demyelinating neurological disease.
 XX
 XX Homo sapiens.
 OS
 XX WO200002917-A2.
 FN
 XX 20-JAN-2000.
 PD
 XX 12-JUL-1999; 99WO-CA000627.
 PF
 XX 10-JUL-1998; 98US-00113977.
 PR
 XX (ADHE-) ADHEREX TECHNOLOGIES INC.
 PA Doherty P, Blaschuk OW, Gour BJ;
 XX WPI; 2000-160899/14.
 DR
 XX Modulating agents for treating autoimmune diseases, cancer, spinal cord
 PT injuries, and for increasing vasopermeability, inhibiting synaptic
 PT stability and preventing pregnancy.
 XX
 XX Disclosure; Fig 2; 144pp; English.
 XX
 XX The present invention describes a cell adhesion modulating agent (I)
 CC capable of binding to the cadherin cell adhesion recognition (CAR)
 CC sequence HAV, where the agent doesn't comprise an antibody or antigen-

CC binding fragment of it. (I) is used for inhibiting or enhancing cadherin
CC mediated functions like cell adhesion, neurite outgrowth, Schwann cell
CC migration and synaptic stability in cells preferably epithelial,
CC endothelial, neural, tumour cells and lymphocytes expressing cadherin E
CC or N. Inhibition of cadherin mediated cell adhesion by (I) is used in
CC reducing unwanted cellular adhesion, enhancing drug delivery through
CC skin, drug delivery to a tumour, treating cancer and/or inhibiting
CC metastasis, inducing apoptosis, inhibiting angiogenesis, modulating
CC immune system, preventing pregnancy, increasing vasopermeability,
CC inhibiting synaptic stability in a mammal. Enhancement is used for
CC facilitating wound healing, enhancing adhesion of foreign tissue implant,
CC enhancing and/or directing neurite outgrowth and treating spinal cord
CC injuries. (I) is used for treating demyelinating neurological diseases
CC and for facilitating blood sampling in a mammal. (I) is used for
CC detecting cadherin expressing cells in a sample by detecting the level of
CC antibody cadherin complex. AA78148 to AA78242 represent sequences used
CC in the exemplification of the present invention
XX
SQ Sequence 108 AA;

Query Match 13.0%; Score 108; DB 3; Length 108;
Best Local Similarity 100.0%; Pred. No. 3.4e-93;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 108 DWVAPISVPENGKGFPPQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWL 167
DB 1 DWVAPISVPENGKGFPPQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWL 60
QY 168 LLNKLDRREIAKYELFGHANSNGASVEDPMNISIIIVTDQNDHKPKF 215
DB 61 LLNKLDRREIAKYELFGHANSNGASVEDPMNISIIIVTDQNDHKPKF 108

Search completed: September 21, 2004, 22:23:04
Job time : 136 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004. Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 22:20:55 : Search time 34 Seconds
(without alignments)
1258.763 Million cell updates/sec

Title: US-09-916-849A-1
Perfect score: 829
Sequence: 1 MGLPRGPLASLLLLQVCWLQ.....NEWGRFKKLADMYGGCEDD 829

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 0
Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/aaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/aaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/aaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/aaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/aaa/3CTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/aaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	108	13.0	108	3	US-08-893-534A-4
2	108	13.0	108	3	US-08-996-679-4
3	108	13.0	108	3	US-08-939-853A-6
4	108	13.0	108	3	US-09-115-395-4
5	108	13.0	108	3	US-09-113-977C-37
6	108	13.0	108	4	US-09-507-102-4
7	108	13.0	108	4	US-09-250-059-4
8	108	13.0	108	4	US-09-248-074-4
9	108	13.0	108	4	US-09-357-717-4
10	108	13.0	108	4	US-09-458-870-4
11	108	13.0	108	4	US-09-351-048A-37
12	108	13.0	108	4	US-09-248-015-4
13	108	13.0	108	4	US-09-544-782-4
14	106	12.8	106	3	US-09-113-977C-46
15	106	12.8	106	4	US-09-351-048A-46
16	51	6.2	822	2	US-08-474-067-9
17	51	6.2	822	2	US-08-474-068A-9
18	51	6.2	822	2	US-08-472-481-8
19	43	5.2	43	4	US-09-057-363C-56
20	43	5.2	43	4	US-09-265-107-56
21	22	2.7	107	1	US-08-326-117B-3
22	22	2.7	107	3	US-08-982-129-3
23	22	2.7	108	3	US-08-893-534A-5
24	22	2.7	108	3	US-08-996-679-5
25	22	2.7	108	3	US-08-939-853A-7
26	22	2.7	108	3	US-09-115-395-5
27	22	2.7	108	3	US-09-113-977C-38

28	22	2.7	108	4	US-09-507-102-5	Sequence 5, Appl
29	22	2.7	108	4	US-09-250-059-5	Sequence 5, Appl
30	22	2.7	108	4	US-09-248-074-5	Sequence 5, Appl
31	22	2.7	108	4	US-09-357-717-5	Sequence 5, Appl
32	22	2.7	108	4	US-09-458-870-5	Sequence 5, Appl
33	22	2.7	108	4	US-09-351-048A-38	Sequence 38, Appl
34	22	2.7	108	4	US-09-248-015-5	Sequence 5, Appl
35	22	2.7	108	4	US-09-544-782-5	Sequence 5, Appl
36	22	2.7	109	4	US-09-178-176B-8	Sequence 8, Appl
37	22	2.7	109	4	US-09-457-864-8	Sequence 8, Appl
38	22	2.7	109	4	US-09-457-865A-8	Sequence 8, Appl
39	17	2.1	837	2	US-08-474-067-7	Sequence 7, Appl
40	17	2.1	837	2	US-08-474-068A-7	Sequence 7, Appl
41	17	2.1	837	2	US-08-472-481-6	Sequence 6, Appl
42	15	1.8	884	2	US-08-474-067-8	Sequence 8, Appl
43	15	1.8	884	2	US-08-474-068A-8	Sequence 8, Appl
44	15	1.8	884	2	US-08-472-481-7	Sequence 7, Appl
45	14	1.7	43	4	US-09-057-363C-57	Sequence 57, Appl

ALIGNMENTS

RESULT 1
US-08-893-534A-4
; Sequence 4, Application US/08893534A
; Patent No. 6031072
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING CELL ADHESION
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/893,534A
; FILING DATE: 11-JUL-1997
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Yaki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 100086.401
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 108 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-893-534A-4

Query Match 13.0%; Score 108; DB 3; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.1e-96;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	108	DWVAPISVPENGKPPQRLNQLKSNKDRDTKIFYITGPGADSPPEGVFAVEKETGWL 167
Db	1	DWVAPISVPENGKPPQRLNQLKSNKDRDTKIFYITGPGADSPPEGVFAVEKETGWL 60
Qy	168	LLNKPLDREETAKYELFGHAUSNGASVEDPMNISIIIVTDNDHKPF 215

Db 61 LLNKPLDREIAKYELFGHVSNGASVEDPNNISIIIVTDQNDHKPKF 108
RESULT 2
US-08-996-679-4
; Sequence 4, Application US/08996679
; Patent No. 6169071
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING
; TITLE OF INVENTION: CELL ADHESION
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 23-DEC-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 100086.401C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 108 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-996-679-4
Query Match 13.0%; Score 108; DB 3; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.1e-96;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 108 DWVAPISVPENGKGFPPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 167
Db 1 DWVAPISVPENGKGFPPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 60
Qy 168 LLNKPLDREIAKYELFGHVSNGASVEDPNNISIIIVTDQNDHKPKF 215
Db 61 LLNKPLDREIAKYELFGHVSNGASVEDPNNISIIIVTDQNDHKPKF 108
RESULT 3
US-08-939-853A-6
; Sequence 6, Application US/08939853A
; Patent No. 6203788
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR REGULATING
; TITLE OF INVENTION: CELL ADHESION
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington

COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/939,853A
FILING DATE: 29-SEP-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 32,391
REFERENCE/DOCKET NUMBER: 100086.402
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 108 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-939-853A-6
Query Match 13.0%; Score 108; DB 3; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.1e-96;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 108 DWVAPISVPENGKGFPPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 167
Db 1 DWVAPISVPENGKGFPPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 60
Qy 168 LLNKPLDREIAKYELFGHVSNGASVEDPNNISIIIVTDQNDHKPKF 215
Db 61 LLNKPLDREIAKYELFGHVSNGASVEDPNNISIIIVTDQNDHKPKF 108
RESULT 4
US-09-115-395-4
; Sequence 4, Application US/09115395A
; Patent No. 6207639
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING NEURITE OUTGROWTH
; FILE REFERENCE: 100086.401C3
; CURRENT APPLICATION NUMBER: US/09/115,395A
; CURRENT FILING DATE: 1998-07-14
; EARLIER APPLICATION NUMBER: 08/996,679
; EARLIER FILING DATE: 1997-12-23
; EARLIER APPLICATION NUMBER: 08/893,534
; EARLIER FILING DATE: 1997-07-11
; EARLIER APPLICATION NUMBER: 60/021,612
; EARLIER FILING DATE: 1996-07-12
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 108
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-115-395-4
Query Match 13.0%; Score 108; DB 3; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.1e-96;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 108 DWVAPISVPENGKGFPPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 167
Db 1 DWVAPISVPENGKGFPPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 60
Qy 168 LLNKPLDREIAKYELFGHVSNGASVEDPNNISIIIVTDQNDHKPKF 215

Db 61 LLNKPLDREETIAKYELFGHVAENGASVEDPMNISIIVTDQNDHKPKF 108
|||||

RESULT 5

US-09-113-977C-37
; Sequence 37, Application US/09113977C
; Patent No. 6277824
; GENERAL INFORMATION:
; APPLICANT: Doherty, Patrick
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING ADHESION MOLECULE
; FILE REFERENCE: 100086.403
; CURRENT APPLICATION NUMBER: US/09/113,977C
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 37
; LENGTH: 108
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-113-977C-37

Query Match 13.0%; Score 108; DB 3; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.1e-96;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 108 DWVAPISVPENGKGPFPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 167
Db 1 DWVAPISVPENGKGPFPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 60
QY 168 LLNKPLDREETIAKYELFGHVAENGASVEDPMNISIIVTDQNDHKPKF 215
Db 61 LLNKPLDREETIAKYELFGHVAENGASVEDPMNISIIVTDQNDHKPKF 108

RESULT 6

US-09-507-102-4
; Sequence 4, Application US/09507102
; Patent No. 6326352
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING CELL ADHESION
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESS: SEED IP LAW GROUP PLLC
; STREET: 6300 Bank of America Bldg., 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/507,102
; FILING DATE: 17-Feb-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/893,534
; FILING DATE: 11-JUL-1997
; APPLICATION NUMBER: US 60/021,612
; FILING DATE: 12-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Christiansen, William T.
; REGISTRATION NUMBER: 44,614
; REFERENCE/DOCKET NUMBER: 100086.401C10

TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 108 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-507-102-4

Query Match 13.0%; Score 108; DB 4; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.1e-96;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 108 DWVAPISVPENGKGPFPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 167
Db 1 DWVAPISVPENGKGPFPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 60
QY 168 LLNKPLDREETIAKYELFGHVAENGASVEDPMNISIIVTDQNDHKPKF 215
Db 61 LLNKPLDREETIAKYELFGHVAENGASVEDPMNISIIVTDQNDHKPKF 108

RESULT 7

US-09-250-059-4
; Sequence 4, Application US/09250059
; Patent No. 6333307
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING NEURITE OUTGROWTH
; FILE REFERENCE: 100086.401C6
; CURRENT APPLICATION NUMBER: US/09/250,059
; CURRENT FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 108
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-250-059-4

Query Match 13.0%; Score 108; DB 4; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.1e-96;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 108 DWVAPISVPENGKGPFPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 167
Db 1 DWVAPISVPENGKGPFPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 60
QY 168 LLNKPLDREETIAKYELFGHVAENGASVEDPMNISIIVTDQNDHKPKF 215
Db 61 LLNKPLDREETIAKYELFGHVAENGASVEDPMNISIIVTDQNDHKPKF 108

RESULT 8

US-09-248-074-4
; Sequence 4, Application US/09248074
; Patent No. 6346512
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING CELL ADHESION
; FILE REFERENCE: 100086.401C5
; CURRENT APPLICATION NUMBER: US/09/248,074
; CURRENT FILING DATE: 1999-02-10
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 108
; TYPE: PRT

```
; ORGANISM: Homo sapiens
US-09-248-074-4
Query Match      13.0%; Score 108; DB 4; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.1e-96;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 108 DWVAPISVPENKGGPPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 167
Db 1 DWVAPISVPENKGGPPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 60

Qy 168 LNKPLDREBIAYELFGHVSNGASVEDPMNISIIITDQNDHKPKF 215
Db 61 LNKPLDREBIAYELFGHVSNGASVEDPMNISIIITDQNDHKPKF 108

RESULT 9
US-09-357-717-4
; Sequence 4, Application US/09357717
; Patent No. 6417325
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; APPLICANT: Farookhi, Riaz
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR CANCER THERAPY
; FILE REFERENCE: 100086.401C7
; CURRENT APPLICATION NUMBER: US/09/357,717
; CURRENT FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 108
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-357-717-4
Query Match      13.0%; Score 108; DB 4; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.1e-96;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 108 DWVAPISVPENKGGPPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 167
Db 1 DWVAPISVPENKGGPPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 60

Qy 168 LNKPLDREBIAYELFGHVSNGASVEDPMNISIIITDQNDHKPKF 215
Db 61 LNKPLDREBIAYELFGHVSNGASVEDPMNISIIITDQNDHKPKF 108

RESULT 10
US-09-458-870-4
; Sequence 4, Application US/09458870
; Patent No. 6465427
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; APPLICANT: Farookhi, Riaz
; APPLICANT: Ali, Ammar
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING CELL ADHESION
; FILE REFERENCE: 100086.401C8
; CURRENT APPLICATION NUMBER: US/09/458,870
; CURRENT FILING DATE: 1999-12-10
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 108
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-458-870-4
Query Match      13.0%; Score 108; DB 4; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.1e-96;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 108 DWVAPISVPENKGGPPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 167
Db 1 DWVAPISVPENKGGPPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 60

Qy 168 LNKPLDREBIAYELFGHVSNGASVEDPMNISIIITDQNDHKPKF 215
Db 61 LNKPLDREBIAYELFGHVSNGASVEDPMNISIIITDQNDHKPKF 108

RESULT 11
US-09-351-048A-37
; Sequence 37, Application US/09351048A
; Patent No. 6472368
; GENERAL INFORMATION:
; APPLICANT: Doherty, Patrick
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING ADHESION MOLECULE
; TITLE OF INVENTION: FUNCTION
; FILE REFERENCE: 100086.403C1
; CURRENT APPLICATION NUMBER: US/09/351,048A
; CURRENT FILING DATE: 1999-07-09
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 37
; LENGTH: 108
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Solid Phase
; OTHER INFORMATION: Synthesis
US-09-351-048A-37
Query Match      13.0%; Score 108; DB 4; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.1e-96;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 108 DWVAPISVPENKGGPPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 167
Db 1 DWVAPISVPENKGGPPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 60

Qy 168 LNKPLDREBIAYELFGHVSNGASVEDPMNISIIITDQNDHKPKF 215
Db 61 LNKPLDREBIAYELFGHVSNGASVEDPMNISIIITDQNDHKPKF 108

RESULT 12
US-09-248-015-4
; Sequence 4, Application US/09248015
; Patent No. 6562786
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING APOPTOSIS
; FILE REFERENCE: 100086.401C4
; CURRENT APPLICATION NUMBER: US/09/248,015
; CURRENT FILING DATE: 1999-02-10
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 108
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-248-015-4
Query Match      13.0%; Score 108; DB 4; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.1e-96;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 108 DWVAPISVPENKGGPPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 167
Db 1 DWVAPISVPENKGGPPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 60
```

QY 168 LNKPLDREIEIAKYELFGHVSNGASVEDPMNISIIIVTDQNDHKPKF 215
Db 61 LNKPLDREIEIAKYELFGHVSNGASVEDPMNISIIIVTDQNDHKPKF 108

RESULT 13

US-09-544-782-4
; Sequence 4, Application US/09544782
; Patent No. 661082
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; APPLICANT: Farookhi, Riaz
; APPLICANT: Ali, Ammar
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING CELL ADHESION
; FILE REFERENCE: 100086.401C9
; CURRENT APPLICATION NUMBER: US/09/544,782
; CURRENT FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 108
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-544-782-4

Query Match 13.0%; Score 108; DB 4; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.1e-96;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 DWVAPISVPENGKGFPPQRLNQLKSNKDRDKIFYSITPGADSPPEGVFAVEKETGWL 167
Db 1 DWVAPISVPENGKGFPPQRLNQLKSNKDRDKIFYSITPGADSPPEGVFAVEKETGWL 60
QY 168 LNKPLDREIEIAKYELFGHVSNGASVEDPMNISIIIVTDQNDHKPKF 215
Db 61 LNKPLDREIEIAKYELFGHVSNGASVEDPMNISIIIVTDQNDHKPKF 108

RESULT 14

US-09-113-977C-46
; Sequence 46, Application US/09113977C
; Patent No. 6277824
; GENERAL INFORMATION:
; APPLICANT: Doherty, Patrick
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING ADHESION MOLECULE
; FILE REFERENCE: 100086.403
; CURRENT APPLICATION NUMBER: US/09/113,977C
; CURRENT FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-113-977C-46

Query Match 12.8%; Score 106; DB 3; Length 106;
Best Local Similarity 100.0%; Pred. No. 9.5e-95;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 441 VPPSKVVEVEGIGTGPVCVYTAEDPDKENQKISYRILRDPAGWLAMPDPSGQVTA VGT 500
Db 1 VPPSKVVEVEGIGTGPVCVYTAEDPDKENQKISYRILRDPAGWLAMPDPSGQVTA VGT 60
QY 501 LDREDEQFVRNNIYEVVWVLMNDNGSPPTTGTGTLTLLTLIDVNDHGP 546
Db 61 LDREDEQFVRNNIYEVVWVLMNDNGSPPTTGTGTLTLLTLIDVNDHGP 106

RESULT 15

US-09-351-048A-46
; Sequence 46, Application US/09351048A
; Patent No. 6472368
; GENERAL INFORMATION:
; APPLICANT: Doherty, Patrick
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING ADHESION MOLECULE
; FILE REFERENCE: 100086.403C1
; CURRENT APPLICATION NUMBER: US/09/351,048A
; CURRENT FILING DATE: 1999-07-09
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Solid Phase
; OTHER INFORMATION: Synthesis
US-09-351-048A-46

Query Match 12.8%; Score 106; DB 4; Length 106;
Best Local Similarity 100.0%; Pred. No. 9.5e-95;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 441 VPPSKVVEVEGIGTGPVCVYTAEDPDKENQKISYRILRDPAGWLAMPDPSGQVTA VGT 500
Db 1 VPPSKVVEVEGIGTGPVCVYTAEDPDKENQKISYRILRDPAGWLAMPDPSGQVTA VGT 60
QY 501 LDREDEQFVRNNIYEVVWVLMNDNGSPPTTGTGTLTLLTLIDVNDHGP 546
Db 61 LDREDEQFVRNNIYEVVWVLMNDNGSPPTTGTGTLTLLTLIDVNDHGP 106

Search completed: September 21, 2004, 22:27:27
Job time : 35 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 22:26:02 ; Search time 141 Seconds

(without alignments)
1888.081 Million cell updates/sec

Title: US-09-916-849A-1

Perfect score: 829

Sequence: 1 MGLPRGLASLLLLQVCMWLQ.....NEWGSRFKLADMYGGGDD 829

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1342398 seqs, 32113274 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1342398

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/FCI_NEW PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/FCIUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	829	100.0	829	9	US-09-905-983-50
2	829	100.0	829	10	US-09-916-849A-1
3	829	100.0	829	10	US-09-916-849A-1
4	829	100.0	829	12	US-09-975-723A-1
5	829	100.0	829	14	US-10-058-270A-126
6	829	100.0	829	14	US-10-174-677-21
7	829	100.0	829	14	US-10-158-123-1
8	829	100.0	829	15	US-10-295-027-783
9	829	100.0	829	15	US-10-295-027-800
10	829	100.0	829	15	US-10-295-027-863
11	829	100.0	829	15	US-10-295-027-896
12	829	100.0	829	15	US-10-295-027-896
13	829	100.0	829	15	US-10-295-027-896
14	829	100.0	829	15	US-10-295-027-896
15	829	100.0	829	15	US-10-295-027-896

16	728	87.8	829	15	US-10-087-080-14	Sequence 14, Appl
17	728	87.8	829	16	US-10-650-112-16	Sequence 16, Appl
18	108	13.0	108	9	US-09-769-145-4	Sequence 4, Appl
19	108	13.0	108	10	US-09-778-026-6	Sequence 6, Appl
20	108	13.0	108	12	US-10-412-701-4	Sequence 4, Appl
21	108	13.0	108	12	US-10-105-008-4	Sequence 4, Appl
22	108	13.0	108	12	US-10-359-546-4	Sequence 4, Appl
23	108	13.0	108	13	US-10-006-982-4	Sequence 4, Appl
24	108	13.0	108	14	US-10-058-821-4	Sequence 4, Appl
25	108	13.0	108	14	US-10-193-653-37	Sequence 37, Appl
26	108	13.0	108	15	US-10-425-557-4	Sequence 4, Appl
27	108	13.0	108	16	US-10-632-678-4	Sequence 4, Appl
28	108	13.0	108	16	US-10-648-854-6	Sequence 46, Appl
29	106	12.8	106	14	US-10-193-653-46	Sequence 56, Appl
30	43	5.2	43	15	US-10-369-226-56	Sequence 3, Appl
31	22	2.7	107	10	US-09-457-865-3	Sequence 5, Appl
32	22	2.7	108	9	US-09-769-145-5	Sequence 7, Appl
33	22	2.7	108	10	US-09-778-026-7	Sequence 5, Appl
34	22	2.7	108	12	US-10-412-701-5	Sequence 5, Appl
35	22	2.7	108	12	US-10-105-008-5	Sequence 5, Appl
36	22	2.7	108	12	US-10-359-546-5	Sequence 5, Appl
37	22	2.7	108	13	US-10-006-982-5	Sequence 5, Appl
38	22	2.7	108	14	US-10-058-821-5	Sequence 38, Appl
39	22	2.7	108	14	US-10-193-653-38	Sequence 5, Appl
40	22	2.7	108	15	US-10-425-557-5	Sequence 5, Appl
41	22	2.7	108	16	US-10-632-678-5	Sequence 7, Appl
42	22	2.7	108	16	US-10-648-854-7	Sequence 8, Appl
43	22	2.7	109	14	US-10-187-088-8	Sequence 4, Appl
44	19	2.3	19	10	US-09-916-849A-4	Sequence 5, Appl
45	17	2.1	17	10	US-09-916-849A-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1

US-09-905-983-50
; Sequence 50, Application US/09905983
; Patent No. US20020045591A1
; GENERAL INFORMATION:
; APPLICANT: Geiger, Benjamin
; APPLICANT: Ben-Ze'ev, Avri
; APPLICANT: Sadot, Einat
; FILE REFERENCE: 01/22326
; TITLE OF INVENTION: METHODS AND THERAPEUTIC COMPOSITIONS FOR TREATING CANCER
; CURRENT APPLICATION NUMBER: US/09/905.983
; CURRENT FILING DATE: 2001-09-28
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-905-983-50

Query Match 100.0%; Score 829; DB 9; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MGLPRGLASLLLLQVCMWLQCAASEPCRAVPRAEVLLEAGGABOEPCQALGKVFMGCPG	60
Db	1	MGLPRGLASLLLLQVCMWLQCAASEPCRAVPRAEVLLEAGGABOEPCQALGKVFMGCPG	60
Qy	61	CEPALFSTNDODFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDWVAPISVPENG	120
Db	61	CEPALFSTNDODFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDWVAPISVPENG	120
Qy	121	KGPPFQRLNQLSKNDRDKIFYSITGPGADSPPEGVFAVEKFTGWLILNKPDLBREIAK	180
Db	121	KGPPFQRLNQLSKNDRDKIFYSITGPGADSPPEGVFAVEKFTGWLILNKPDLBREIAK	180
Qy	181	VELFGHASENGASVEDPMNISILVTDNDHKPKFTQDTFEGSVLEGLVPGTSVMQVAT	240

```
Db 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTRFGSVLEGVLGTSVMQVAT 240
Qy 241 DEDDAIYTYNGVAVSYHSQEPKDPHLMFTIHRSTGTISVSSGLDREKVPYTLTIOA 300
Db 241 DEDDAIYTYNGVAVSYHSQEPKDPHLMFTIHRSTGTISVSSGLDREKVPYTLTIOA 300
Qy 301 TDMGDGSTTTAVAVVEILDANDNAPMFDPOKYEAHVPENAVGHEVQRLTVTDLDAENSP 360
Db 301 TDMGDGSTTTAVAVVEILDANDNAPMFDPOKYEAHVPENAVGHEVQRLTVTDLDAENSP 360
Qy 361 AWRATYILMGDDGDGHFTITTHPESNGILTRKGLDFAKNQHTLYVEVTNEAPFVLKL 420
Db 361 AWRATYILMGDDGDGHFTITTHPESNGILTRKGLDFAKNQHTLYVEVTNEAPFVLKL 420
Qy 421 PTSTATIVVHVEDVNEAPVFPVPSKVVEQEGIPTGEPVCVYTAEDPDKENQKISYRILR 480
Db 421 PTSTATIVVHVEDVNEAPVFPVPSKVVEQEGIPTGEPVCVYTAEDPDKENQKISYRILR 480
Qy 481 DPAGLWAMPDSGOVTAVGTLDRDEQFVRNNIYEVVWVLMNDNGSPPTTGTGLLLTLID 540
Db 481 DPAGLWAMPDSGOVTAVGTLDRDEQFVRNNIYEVVWVLMNDNGSPPTTGTGLLLTLID 540
Qy 541 VNDHGPVPEPRQITICNQSPPRVHLNITDKLSPHTSPFOAQLTDDSDIYWTAEVNEEGD 600
Db 541 VNDHGPVPEPRQITICNQSPPRVHLNITDKLSPHTSPFOAQLTDDSDIYWTAEVNEEGD 600
Qy 601 TVVLSLKKFLKQDHYDHLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFILPVL 660
Db 601 TVVLSLKKFLKQDHYDHLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFILPVL 660
Qy 661 GAVLALLFLLVLLVLRKKRKEPILLPEDDTRDNVFFYBEGGEEEDQDIDITQLHR 720
Db 661 GAVLALLFLLVLLVLRKKRKEPILLPEDDTRDNVFFYBEGGEEEDQDIDITQLHR 720
Qy 721 GLEARPEVLRNDVAPTIPTMYRPRPNDPDEIGNFIENLKAANTDPTAPPYDILLVF 780
Db 721 GLEARPEVLRNDVAPTIPTMYRPRPNDPDEIGNFIENLKAANTDPTAPPYDILLVF 780
Qy 781 DYEGSGDAASLSLTSASDQDDYDLNWSRPFKKLADMYGGGEDD 829
Db 781 DYEGSGDAASLSLTSASDQDDYDLNWSRPFKKLADMYGGGEDD 829

RESULT 2
US-09-916-849A-1
; Sequence 1, Application US/09916849A
; Publication No. US20030086934A1
; GENERAL INFORMATION:
; APPLICANT: Bostein, et al.
; TITLE OF INVENTION: Basal Markers in Breast Cancer and Related Reagents
; FILE REFERENCE: 2002850-0024
; CURRENT APPLICATION NUMBER: US/09/916,849A
; CURRENT FILING DATE: 2001-07-26
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 1
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Sequence of
; OTHER INFORMATION: Cadherin 3
US-09-916-849A-1

Query Match 100.0%; Score 829; DB 10; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGLPRGLASLLQLVCVILQCAASEPCRAVFRAEVTLFAGGAEQEPGALGVFMGCPG 60
Db 1 MGLPRGLASLLQLVCVILQCAASEPCRAVFRAEVTLFAGGAEQEPGALGVFMGCPG 60
```

```
Qy 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDVWVAPISVPENG 120
Db 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDVWVAPISVPENG 120
Qy 121 KGPFPQRLNQKSKNRDRTKIFYSITGPADSPPEGVFAVEKETGWLKLLNKLPLDREBIAK 180
Db 121 KGPFPQRLNQKSKNRDRTKIFYSITGPADSPPEGVFAVEKETGWLKLLNKLPLDREBIAK 180
Qy 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTRFGSVLEGVLGTSVMQVAT 240
Db 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTRFGSVLEGVLGTSVMQVAT 240
Qy 241 DEDDAIYTYNGVAVSYHSQEPKDPHLMFTIHRSTGTISVSSGLDREKVPYTLTIOA 300
Db 241 DEDDAIYTYNGVAVSYHSQEPKDPHLMFTIHRSTGTISVSSGLDREKVPYTLTIOA 300
Qy 301 TDMGDGSTTTAVAVVEILDANDNAPMFDPOKYEAHVPENAVGHEVQRLTVTDLDAENSP 360
Db 301 TDMGDGSTTTAVAVVEILDANDNAPMFDPOKYEAHVPENAVGHEVQRLTVTDLDAENSP 360
Qy 361 AWRATYILMGDDGDGHFTITTHPESNGILTRKGLDFAKNQHTLYVEVTNEAPFVLKL 420
Db 361 AWRATYILMGDDGDGHFTITTHPESNGILTRKGLDFAKNQHTLYVEVTNEAPFVLKL 420
Qy 421 PTSTATIVVHVEDVNEAPVFPVPSKVVEQEGIPTGEPVCVYTAEDPDKENQKISYRILR 480
Db 421 PTSTATIVVHVEDVNEAPVFPVPSKVVEQEGIPTGEPVCVYTAEDPDKENQKISYRILR 480
Qy 481 DPAGLWAMPDSGOVTAVGTLDRDEQFVRNNIYEVVWVLMNDNGSPPTTGTGLLLTLID 540
Db 481 DPAGLWAMPDSGOVTAVGTLDRDEQFVRNNIYEVVWVLMNDNGSPPTTGTGLLLTLID 540
Qy 541 VNDHGPVPEPRQITICNQSPPRVHLNITDKLSPHTSPFOAQLTDDSDIYWTAEVNEEGD 600
Db 541 VNDHGPVPEPRQITICNQSPPRVHLNITDKLSPHTSPFOAQLTDDSDIYWTAEVNEEGD 600
Qy 601 TVVLSLKKFLKQDHYDHLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFILPVL 660
Db 601 TVVLSLKKFLKQDHYDHLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFILPVL 660
Qy 661 GAVLALLFLLVLLVLRKKRKEPILLPEDDTRDNVFFYBEGGEEEDQDIDITQLHR 720
Db 661 GAVLALLFLLVLLVLRKKRKEPILLPEDDTRDNVFFYBEGGEEEDQDIDITQLHR 720
Qy 721 GLEARPEVLRNDVAPTIPTMYRPRPNDPDEIGNFIENLKAANTDPTAPPYDILLVF 780
Db 721 GLEARPEVLRNDVAPTIPTMYRPRPNDPDEIGNFIENLKAANTDPTAPPYDILLVF 780
Qy 781 DYEGSGDAASLSLTSASDQDDYDLNWSRPFKKLADMYGGGEDD 829
Db 781 DYEGSGDAASLSLTSASDQDDYDLNWSRPFKKLADMYGGGEDD 829

RESULT 3
US-09-975-723A-1
; Sequence 1, Application US/09975723A
; Publication No. US20030108529A1
; GENERAL INFORMATION:
; APPLICANT: Nackman, Gary
; APPLICANT: Poty, Ramsey
; TITLE OF INVENTION: Improvement of Endothelial Cell-Cell
; TITLE OF INVENTION: Cohesion
; FILE REFERENCE: 601-1-101N
; CURRENT APPLICATION NUMBER: US/09/975,723A
; CURRENT FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: 60/241,216
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/243,693
; PRIOR FILING DATE: 2000-10-27
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 829
```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-975-723A-1

Query Match      100.0%; Score 829; DB 10; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPRGPLASLLLLQVCWLCQAASEPCRAVFREAEVTLLEAGGAEQEPGQALGKVFMGCPG 60
DB 1 MGLPRGPLASLLLLQVCWLCQAASEPCRAVFREAEVTLLEAGGAEQEPGQALGKVFMGCPG 60
QY 61 QEPALFSTDNDDFTVRNGETVQERSLKERNPLKIPPSKRIILRRHKRDVVAPISVPENG 120
DB 61 QEPALFSTDNDDFTVRNGETVQERSLKERNPLKIPPSKRIILRRHKRDVVAPISVPENG 120
QY 121 KGPPFORNLQKSNKDRDTKIFYITGPGADSPPEGVFAVEKETGMLLNKPLDREBIAK 180
DB 121 KGPPFORNLQKSNKDRDTKIFYITGPGADSPPEGVFAVEKETGMLLNKPLDREBIAK 180
QY 181 YELFGHAVSENGASVEDPNNISIIYTDQNDHKPKETQDTFRGSVLEGLVPGTSVMQVAT 240
DB 181 YELFGHAVSENGASVEDPNNISIIYTDQNDHKPKETQDTFRGSVLEGLVPGTSVMQVAT 240
QY 241 DEDDAIYTYNGVAVYSIHQEPKDPHDLMTIHRSTGTISVSSGLDREKVEYTLTIQA 300
DB 241 DEDDAIYTYNGVAVYSIHQEPKDPHDLMTIHRSTGTISVSSGLDREKVEYTLTIQA 300
QY 301 TMDGDSGTTTAVAVVEILDANDNAPMFPDQKYEAHVPENAVGHEVORLTVDLDAPNSP 360
DB 301 TMDGDSGTTTAVAVVEILDANDNAPMFPDQKYEAHVPENAVGHEVORLTVDLDAPNSP 360
QY 361 AWRATVILMGDDGDHFTITTHPESNOGILTRKGLDFAKNOHTLYVEVNEAPFVLKL 420
DB 361 AWRATVILMGDDGDHFTITTHPESNOGILTRKGLDFAKNOHTLYVEVNEAPFVLKL 420
QY 421 PTSTATIVVHVEDVNEAPVFVPPSKVVEQSGIPTGEPVCVYTABDPDKENOKISYRIIL 480
DB 421 PTSTATIVVHVEDVNEAPVFVPPSKVVEQSGIPTGEPVCVYTABDPDKENOKISYRIIL 480
QY 481 DPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVWVLMNDGSPPTGTGTLTLID 540
DB 481 DPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVWVLMNDGSPPTGTGTLTLID 540
QY 541 VNDHGVPVPRQITICNQSPVRHLNITDKLSPTSPPFQAQLTDDSDIYWTAEVNEEGD 600
DB 541 VNDHGVPVPRQITICNQSPVRHLNITDKLSPTSPPFQAQLTDDSDIYWTAEVNEEGD 600
QY 601 TVVLSLKKFLKQDITYDVHLSLDHGNKEQITVIRATVCDCHGHVETCPGWKGGFILPVL 660
DB 601 TVVLSLKKFLKQDITYDVHLSLDHGNKEQITVIRATVCDCHGHVETCPGWKGGFILPVL 660
QY 661 GAVIALLFLLVLLLVRRKRKIKEPLLPEDDTRDNVFFYEGEGGEDDQDYDITQLHR 720
DB 661 GAVIALLFLLVLLLVRRKRKIKEPLLPEDDTRDNVFFYEGEGGEDDQDYDITQLHR 720
QY 721 GLEAREVVLNDVAPTIPTMYRRPRANPBEIGNFIENIKAANTDPTAPPDTLLVF 780
DB 721 GLEAREVVLNDVAPTIPTMYRRPRANPBEIGNFIENIKAANTDPTAPPDTLLVF 780
QY 781 DYEGSGDAASLSLTSASDQDQDYDLNWSGRFVKLLADMYGGGEDD 829
DB 781 DYEGSGDAASLSLTSASDQDQDYDLNWSGRFVKLLADMYGGGEDD 829
```

RESULT 4

```
US-10-058-270A-126
; Sequence 126, Application US/10058270A
; Publication No. US20040029114A1
; GENERAL INFORMATION:
; APPLICANT: Mack, David H.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Afar, Daniel
```

```
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of diagnosis of Breast Cancer, Compositions and
; FILE OF INVENTION: Methods of Screening for Modulators of Breast Cancer
; FILE REFERENCE: 018501-005210US
; CURRENT APPLICATION NUMBER: US/10/058,270A
; CURRENT FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: US 60/263,965
; PRIOR FILING DATE: 2001-01-24
; PRIOR APPLICATION NUMBER: US 60/265,928
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: US 09/829,472
; PRIOR FILING DATE: 2001-04-09
; PRIOR APPLICATION NUMBER: US 60/282,698
; PRIOR FILING DATE: 2001-04-09
; PRIOR APPLICATION NUMBER: US 60/288,590
; PRIOR FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: US 60/294,443
; PRIOR FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 141
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 126
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-058-270A-126
```

```
Query Match      100.0%; Score 829; DB 12; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPRGPLASLLLLQVCWLCQAASEPCRAVFREAEVTLLEAGGAEQEPGQALGKVFMGCPG 60
DB 1 MGLPRGPLASLLLLQVCWLCQAASEPCRAVFREAEVTLLEAGGAEQEPGQALGKVFMGCPG 60
QY 61 QEPALFSTDNDDFTVRNGETVQERSLKERNPLKIPPSKRIILRRHKRDVVAPISVPENG 120
DB 61 QEPALFSTDNDDFTVRNGETVQERSLKERNPLKIPPSKRIILRRHKRDVVAPISVPENG 120
QY 121 KGPPFORNLQKSNKDRDTKIFYITGPGADSPPEGVFAVEKETGMLLNKPLDREBIAK 180
DB 121 KGPPFORNLQKSNKDRDTKIFYITGPGADSPPEGVFAVEKETGMLLNKPLDREBIAK 180
QY 181 YELFGHAVSENGASVEDPNNISIIYTDQNDHKPKETQDTFRGSVLEGLVPGTSVMQVAT 240
DB 181 YELFGHAVSENGASVEDPNNISIIYTDQNDHKPKETQDTFRGSVLEGLVPGTSVMQVAT 240
QY 241 DEDDAIYTYNGVAVYSIHQEPKDPHDLMTIHRSTGTISVSSGLDREKVEYTLTIQA 300
DB 241 DEDDAIYTYNGVAVYSIHQEPKDPHDLMTIHRSTGTISVSSGLDREKVEYTLTIQA 300
QY 301 TMDGDSGTTTAVAVVEILDANDNAPMFPDQKYEAHVPENAVGHEVORLTVDLDAPNSP 360
DB 301 TMDGDSGTTTAVAVVEILDANDNAPMFPDQKYEAHVPENAVGHEVORLTVDLDAPNSP 360
QY 361 AWRATVILMGDDGDHFTITTHPESNOGILTRKGLDFAKNOHTLYVEVNEAPFVLKL 420
DB 361 AWRATVILMGDDGDHFTITTHPESNOGILTRKGLDFAKNOHTLYVEVNEAPFVLKL 420
QY 421 PTSTATIVVHVEDVNEAPVFVPPSKVVEQSGIPTGEPVCVYTABDPDKENOKISYRIIL 480
DB 421 PTSTATIVVHVEDVNEAPVFVPPSKVVEQSGIPTGEPVCVYTABDPDKENOKISYRIIL 480
QY 481 DPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVWVLMNDGSPPTGTGTLTLID 540
DB 481 DPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVWVLMNDGSPPTGTGTLTLID 540
QY 541 VNDHGVPVPRQITICNQSPVRHLNITDKLSPTSPPFQAQLTDDSDIYWTAEVNEEGD 600
DB 541 VNDHGVPVPRQITICNQSPVRHLNITDKLSPTSPPFQAQLTDDSDIYWTAEVNEEGD 600
QY 601 TVVLSLKKFLKQDITYDVHLSLDHGNKEQITVIRATVCDCHGHVETCPGWKGGFILPVL 660
DB 601 TVVLSLKKFLKQDITYDVHLSLDHGNKEQITVIRATVCDCHGHVETCPGWKGGFILPVL 660
```


QY 661 GAVLALLFLLLVLLVLLVLRKKRKEIPEDDTRDNVYFYGEGGEEDQDYDITQLHR 720
Db 661 GAVLALLFLLLVLLVLLVLRKKRKEIPEDDTRDNVYFYGEGGEEDQDYDITQLHR 720
QY 721 GLEARPEVLNVDVAPITIIPTMYRPRPANDPDSIGNFIENKAANTDTPAPPYDITLLVF 780
Db 721 GLEARPEVLNVDVAPITIIPTMYRPRPANDPDSIGNFIENKAANTDTPAPPYDITLLVF 780
QY 781 DYEGSGDAASLSLTSASDQDDYDLNENWSRPFKKLADMYGGGEDD 829
Db 781 DYEGSGDAASLSLTSASDQDDYDLNENWSRPFKKLADMYGGGEDD 829
RESULT 5
US-10-174-677-21
; Sequence 21, Application US/10174677
; Publication No. US20030190704A1
; GENERAL INFORMATION:
; APPLICANT: Xie, Ting
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ANCHORING STEM CELLS IN A MICROENVIR
; FILE REFERENCE: 40716(IP-012)
; CURRENT APPLICATION NUMBER: US/10/174,677
; CURRENT FILING DATE: 2002-06-19
; NUMBER OF SEQ ID NOS: 117
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 21
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-174-677-21
Query Match 100.0%; Score 829; DB 14; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGLPRGPLASLLLLQVCWLCQAASEPCRAVFREAEVTLRAGAEQBPQALGKVFVWGCPG 60
Db 1 MGLPRGPLASLLLLQVCWLCQAASEPCRAVFREAEVTLRAGAEQBPQALGKVFVWGCPG 60
QY 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIPPSKRILRRHRKRDWVAPISVPENG 120
Db 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIPPSKRILRRHRKRDWVAPISVPENG 120
QY 121 KGFPPLNQLKSKNDKDTKIFYSITGPGADSPPEGVFAVEKETGWLILNKLPLDREELAK 180
Db 121 KGFPPLNQLKSKNDKDTKIFYSITGPGADSPPEGVFAVEKETGWLILNKLPLDREELAK 180
QY 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSVLEGLVPGTSMQVAT 240
Db 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSVLEGLVPGTSMQVAT 240
QY 241 DEDDAIYTVNGVAVSIHSGEPKDPHDLMTIHRSTGTISVSSGLDREKVEPYTLTIOA 300
Db 241 DEDDAIYTVNGVAVSIHSGEPKDPHDLMTIHRSTGTISVSSGLDREKVEPYTLTIOA 300
QY 301 TMDGSGSTTTAAVAVVEILDANDNAPMFPDQKYEAHVAVPENAVGHEVQRLTVDLDAPNSP 360
Db 301 TMDGSGSTTTAAVAVVEILDANDNAPMFPDQKYEAHVAVPENAVGHEVQRLTVDLDAPNSP 360
QY 361 AWRATYILMGDDGDHFTITTHPESNOGILTRKGLDPEAKNQHTLYEVETNEAPFVLKL 420
Db 361 AWRATYILMGDDGDHFTITTHPESNOGILTRKGLDPEAKNQHTLYEVETNEAPFVLKL 420
QY 421 PTSTAIIVHVEDVNEAPVFPVPSKVVEQEGPTGEPVCVYTAEDDPKENGKISYRILR 480
Db 421 PTSTAIIVHVEDVNEAPVFPVPSKVVEQEGPTGEPVCVYTAEDDPKENGKISYRILR 480
QY 481 DPAGWLAMPDSQGVAVGTLDREDEQFVRNNIYEVWMLAMNGSPPTGTGTLTLTID 540
Db 481 DPAGWLAMPDSQGVAVGTLDREDEQFVRNNIYEVWMLAMNGSPPTGTGTLTLTID 540
QY 541 VNDHGVPPEPRQITICNQSPVRHVLNITDKDLSPHTSPPQAQLTDDSDIYWTAEVNEGSD 600

Db 541 VNDHGVPPEPRQITICNQSPVRHVLNITDKDLSPHTSPPQAQLTDDSDIYWTAEVNEGSD 600
QY 601 TVVLSLKKFLKQDYYDHLVLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKGFIPLVL 660
Db 601 TVVLSLKKFLKQDYYDHLVLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKGFIPLVL 660
QY 661 GAVLALLFLLLVLLVLLVLRKKRKEIPEDDTRDNVYFYGEGGEEDQDYDITQLHR 720
Db 661 GAVLALLFLLLVLLVLLVLRKKRKEIPEDDTRDNVYFYGEGGEEDQDYDITQLHR 720
QY 721 GLEARPEVLNVDVAPITIIPTMYRPRPANDPDSIGNFIENKAANTDTPAPPYDITLLVF 780
Db 721 GLEARPEVLNVDVAPITIIPTMYRPRPANDPDSIGNFIENKAANTDTPAPPYDITLLVF 780
QY 781 DYEGSGDAASLSLTSASDQDDYDLNENWSRPFKKLADMYGGGEDD 829
Db 781 DYEGSGDAASLSLTSASDQDDYDLNENWSRPFKKLADMYGGGEDD 829
RESULT 6
US-10-158-123-1
; Sequence 1, Application US/10158123
; Publication No. US20030194406A1
; GENERAL INFORMATION:
; APPLICANT: Reinhard, Christoph
; APPLICANT: Klinger, Julie
; APPLICANT: Jefferson, Ann
; APPLICANT: Escobedo, Jaime
; APPLICANT: Randazzo, Fillipo
; APPLICANT: Winter, Jill
; APPLICANT: Goodson, Robert
; APPLICANT: Qi, Weimin
; TITLE OF INVENTION: P-Cadherin as a Target for Anti-Cancer
; FILE REFERENCE: 35784/258994
; CURRENT APPLICATION NUMBER: US/10/158,123
; CURRENT FILING DATE: 2003-02-07
; PRIOR APPLICATION NUMBER: 60/294,225
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-158-123-1
Query Match 100.0%; Score 829; DB 14; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGLPRGPLASLLLLQVCWLCQAASEPCRAVFREAEVTLRAGAEQBPQALGKVFVWGCPG 60
Db 1 MGLPRGPLASLLLLQVCWLCQAASEPCRAVFREAEVTLRAGAEQBPQALGKVFVWGCPG 60
QY 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIPPSKRILRRHRKRDWVAPISVPENG 120
Db 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIPPSKRILRRHRKRDWVAPISVPENG 120
QY 121 KGFPPLNQLKSKNDKDTKIFYSITGPGADSPPEGVFAVEKETGWLILNKLPLDREELAK 180
Db 121 KGFPPLNQLKSKNDKDTKIFYSITGPGADSPPEGVFAVEKETGWLILNKLPLDREELAK 180
QY 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSVLEGLVPGTSMQVAT 240
Db 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSVLEGLVPGTSMQVAT 240
QY 241 DEDDAIYTVNGVAVSIHSGEPKDPHDLMTIHRSTGTISVSSGLDREKVEPYTLTIOA 300
Db 241 DEDDAIYTVNGVAVSIHSGEPKDPHDLMTIHRSTGTISVSSGLDREKVEPYTLTIOA 300
QY 301 TMDGSGSTTTAAVAVVEILDANDNAPMFPDQKYEAHVAVPENAVGHEVQRLTVDLDAPNSP 360

RESULT 8
US-10-295-027-800
; Sequence 800, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 800
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-800

Query Match 100.0%; Score 829; DB 15; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MGLPRGLASLLLLQVCWLQCAASEPCRAVFREAEVTLSEAGAEQEPGALGKVFMGCPG	60
DB	1	MGLPRGLASLLLLQVCWLQCAASEPCRAVFREAEVTLSEAGAEQEPGALGKVFMGCPG	60
QY	61	QEPALFSTNDFFTNGETVQERRSLKERNPLKIPPSKRIILRRHKRWVAPISPENG	120
DB	61	QEPALFSTNDFFTNGETVQERRSLKERNPLKIPPSKRIILRRHKRWVAPISPENG	120
QY	121	KGFPQRLNQLSKNDKRTKIFYSITPGADSPPEGVFAVEKETGWLINLKLPLDREBIAK	180
DB	121	KGFPQRLNQLSKNDKRTKIFYSITPGADSPPEGVFAVEKETGWLINLKLPLDREBIAK	180
QY	181	YELFGHVGASGVEDPNISIIYTDQNDHKPKFTQDTFRGSVLEGLPGTSVMQVTAT	240
DB	181	YELFGHVGASGVEDPNISIIYTDQNDHKPKFTQDTFRGSVLEGLPGTSVMQVTAT	240
QY	241	DEDDAIYTVNGVAVYIHSQEPKDPHDLMTFTIHRSTGTISVISGLDREKVPYTLTIOA	300
DB	241	DEDDAIYTVNGVAVYIHSQEPKDPHDLMTFTIHRSTGTISVISGLDREKVPYTLTIOA	300

QY	301	TDMDGSGTTTAVAVVEILDANDANPMPDPKPYEAHPVENAVGHEVORLTVTDLDAENSP	360
DB	301	TDMDGSGTTTAVAVVEILDANDANPMPDPKPYEAHPVENAVGHEVORLTVTDLDAENSP	360
QY	361	AWRATYILMGDDGDHFTITTHPESNQIILTRKGLDPEAKNQHTLYVEVTNEAPFVLKL	420
DB	361	AWRATYILMGDDGDHFTITTHPESNQIILTRKGLDPEAKNQHTLYVEVTNEAPFVLKL	420
QY	421	PTSTATIVVHVEDVNEAPVFPVPSKVVEQEGITGPVPCVYTAEDDPKKNOKISYELR	480
DB	421	PTSTATIVVHVEDVNEAPVFPVPSKVVEQEGITGPVPCVYTAEDDPKKNOKISYELR	480
QY	481	DPAGWLAMPDPSGGVTAAGTLDREDEQFVRNNIYEVNVLAMDNGSPPTTGTLLTLID	540
DB	481	DPAGWLAMPDPSGGVTAAGTLDREDEQFVRNNIYEVNVLAMDNGSPPTTGTLLTLID	540
QY	541	VNDHGPVPEPQITICNQSPVRHVLTITDKLSHTSPFOAQLTDDSDIYTAVNEEGD	600
DB	541	VNDHGPVPEPQITICNQSPVRHVLTITDKLSHTSPFOAQLTDDSDIYTAVNEEGD	600
QY	601	TWVLSLKKFLKQDITYDVLHLSLSDHGNKEQLTVIRATVCDCHGVETCPGPKWGGFILPVL	660
DB	601	TWVLSLKKFLKQDITYDVLHLSLSDHGNKEQLTVIRATVCDCHGVETCPGPKWGGFILPVL	660
QY	661	GAVLALLFLLLVLLLVKRRKIKPELLLPEDDTRDNVYFYGGEGGEEODYDITQHR	720
DB	661	GAVLALLFLLLVLLLVKRRKIKPELLLPEDDTRDNVYFYGGEGGEEODYDITQHR	720
QY	721	GLEARPEVLRNDVAPTIITPMYRPRPANDPDEIGNFIENLKAANTDPTAPPYDTLLVF	780
DB	721	GLEARPEVLRNDVAPTIITPMYRPRPANDPDEIGNFIENLKAANTDPTAPPYDTLLVF	780
QY	781	DYEGSGSDAASLSLTSASDQDDYDYLNEGSRFKKLADMYGGGSD 829	
DB	781	DYEGSGSDAASLSLTSASDQDDYDYLNEGSRFKKLADMYGGGSD 829	

RESULT 9
US-10-295-027-863
; Sequence 863, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10

;; PRIOR APPLICATION NUMBER: US 60/355,250
;; PRIOR FILING DATE: 2002-02-08
;; PRIOR APPLICATION NUMBER: US 60/356,714
;; PRIOR FILING DATE: 2002-02-13
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 1386
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 863
;; LENGTH: 829
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-295-027-863

Query Match 100.0%; Score 829; DB 15; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPRGLASLLLLQVCLQCAASEPCRAVREAEVTLLEAGGAEQPGQALGKVFMCPCG 60
DB 1 MGLPRGLASLLLLQVCLQCAASEPCRAVREAEVTLLEAGGAEQPGQALGKVFMCPCG 60

QY 61 QEPALFSTDNDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDWVAVISVPENG 120
DB 61 QEPALFSTDNDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDWVAVISVPENG 120

QY 121 KGFPQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWLILNKPDLREBIK 180
DB 121 KGFPQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWLILNKPDLREBIK 180

QY 181 YELFGHAVSENGASVEDPMNISIIVTDQNDHKPKFTQDTFRGSVLEGLVPGTSVWQVAT 240
DB 181 YELFGHAVSENGASVEDPMNISIIVTDQNDHKPKFTQDTFRGSVLEGLVPGTSVWQVAT 240

QY 241 DEDDAIYTYNGVAVSIHSEQPKDPHLMFTIHRSTGTISVISSGLDREKVPETTLTIOA 300
DB 241 DEDDAIYTYNGVAVSIHSEQPKDPHLMFTIHRSTGTISVISSGLDREKVPETTLTIOA 300

RESULT 10
US-10-295-027-896
;; Sequence 896, Application US/10295027
;; Publication No. US20030232350A1
;; GENERAL INFORMATION:
;; APPLICANT: Afar, Daniel
;; APPLICANT: Aziz, Natasha
;; APPLICANT: Ginsberg, Wendy M.
;; APPLICANT: Gish, Kurt C.
;; APPLICANT: Glynn, Richard
;; APPLICANT: Hevezi, Peter A.
;; APPLICANT: Mack, David H.
;; APPLICANT: Murray, Richard
;; APPLICANT: Watson, Susan R.
;; APPLICANT: Eos Biotechnology, Inc.
;; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
;; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
;; FILE REFERENCE: 018501-012500US
;; CURRENT APPLICATION NUMBER: US/10/295,027
;; CURRENT FILING DATE: 2002-11-13
;; PRIOR APPLICATION NUMBER: US 09/663,733
;; PRIOR FILING DATE: 2000-09-15
;; PRIOR APPLICATION NUMBER: US 60/350,666
;; PRIOR FILING DATE: 2001-11-13
;; PRIOR APPLICATION NUMBER: US 60/335,394
;; PRIOR FILING DATE: 2001-11-15
;; PRIOR APPLICATION NUMBER: US 60/332,464
;; PRIOR FILING DATE: 2001-11-21
;; PRIOR APPLICATION NUMBER: US 60/334,393
;; PRIOR FILING DATE: 2001-11-29
;; PRIOR APPLICATION NUMBER: US 60/340,376
;; PRIOR FILING DATE: 2001-12-14
;; PRIOR APPLICATION NUMBER: US 60/347,211
;; PRIOR FILING DATE: 2002-01-08
;; PRIOR APPLICATION NUMBER: US 60/347,349
;; PRIOR FILING DATE: 2002-01-10
;; PRIOR APPLICATION NUMBER: US 60/355,250
;; PRIOR FILING DATE: 2002-02-08
;; PRIOR APPLICATION NUMBER: US 60/356,714
;; PRIOR FILING DATE: 2002-02-13
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 1386
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 896
;; LENGTH: 829
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-295-027-896

Query Match 100.0%; Score 829; DB 15; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPRGLASLLLLQVCLQCAASEPCRAVREAEVTLLEAGGAEQPGQALGKVFMCPCG 60
DB 1 MGLPRGLASLLLLQVCLQCAASEPCRAVREAEVTLLEAGGAEQPGQALGKVFMCPCG 60

QY 61 QEPALFSTDNDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDWVAVISVPENG 120
DB 61 QEPALFSTDNDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDWVAVISVPENG 120

QY 121 KGFPQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWLILNKPDLREBIK 180
DB 121 KGFPQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWLILNKPDLREBIK 180

QY 181 YELFGHAVSENGASVEDPMNISIIVTDQNDHKPKFTQDTFRGSVLEGLVPGTSVWQVAT 240
DB 181 YELFGHAVSENGASVEDPMNISIIVTDQNDHKPKFTQDTFRGSVLEGLVPGTSVWQVAT 240

QY 241 DEDDAIYTYNGVAVSIHSEQPKDPHLMFTIHRSTGTISVISSGLDREKVPETTLTIOA 300
DB 241 DEDDAIYTYNGVAVSIHSEQPKDPHLMFTIHRSTGTISVISSGLDREKVPETTLTIOA 300

QY 301 TMDGDSGTTTAVAVVEILDANDNAPMFDPOKYEAHVPENAVGHVORLTVTDLDPNSP 360
Db 301 TMDGDSGTTTAVAVVEILDANDNAPMFDPOKYEAHVPENAVGHVORLTVTDLDPNSP 360
QY 361 AWRATYILMGDDGDHDTITTHPESNOGILTTTRKGLDPEAKNOHTLYVEVNEAPVVLK 420
Db 361 AWRATYILMGDDGDHDTITTHPESNOGILTTTRKGLDPEAKNOHTLYVEVNEAPVVLK 420
QY 421 PTSTATIVHVEDVNEAPVFPVPSKVVEQEGIPTGEPVCVYTAEDPDKENQISYRILR 480
Db 421 PTSTATIVHVEDVNEAPVFPVPSKVVEQEGIPTGEPVCVYTAEDPDKENQISYRILR 480
QY 481 DPAGWLAMPDSGOVTAVGLTDREDEQFVRNNIYEVNVLAMDNQSPPTTGTTGLLLTLD 540
Db 481 DPAGWLAMPDSGOVTAVGLTDREDEQFVRNNIYEVNVLAMDNQSPPTTGTTGLLLTLD 540
QY 541 VNRHGPVPEPRQITICNOSPVRHVLNITDKLSPHTSPFQAQLTDDSDIYWTAEVNEEGD 600
Db 541 VNRHGPVPEPRQITICNOSPVRHVLNITDKLSPHTSPFQAQLTDDSDIYWTAEVNEEGD 600
QY 601 TVVLSLKKFLKQDQYDVHLSLSDHGKKEQLTVIRATVCDCHGVETCPGPKWGGFLLPVL 660
Db 601 TVVLSLKKFLKQDQYDVHLSLSDHGKKEQLTVIRATVCDCHGVETCPGPKWGGFLLPVL 660
QY 661 GAVLALLFLLVLLVLLVLRKKRKEPLLLPEDDTRDNVYFVYEGEGEEDQDYDITQLHR 720
Db 661 GAVLALLFLLVLLVLLVLRKKRKEPLLLPEDDTRDNVYFVYEGEGEEDQDYDITQLHR 720
QY 721 GLEARPEVLRNDVAPTIITPMYRPRPANDPDEIGNFIENLKAANTDTPAPPYDILLVF 780
Db 721 GLEARPEVLRNDVAPTIITPMYRPRPANDPDEIGNFIENLKAANTDTPAPPYDILLVF 780
QY 781 DYEGSGDAASLSLSSASDQDDYDYLNEWGRPKKLADMYGGEDD 829
Db 781 DYEGSGDAASLSLSSASDQDDYDYLNEWGRPKKLADMYGGEDD 829

RESULT 11
US-10-229-345-16
; Sequence 16, Application US/10229345
; Publication No. US20040038220A1
; GENERAL INFORMATION:
; APPLICANT: MARKOWITZ, Sanford D.
; TITLE OF INVENTION: METHODS FOR CATEGORIZING PATIENTS
; FILE REFERENCE: CWRU-P01-003
; CURRENT APPLICATION NUMBER: US/10/229,345
; CURRENT FILING DATE: 2002-08-26
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 16
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-229-345-16

Query Match 87.8%; Score 728; DB 12; Length 829;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 828; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGLPRGLASLLLLQVCWLQCAASEPCRAVFREAEVTLLEAGGAEQEPGALGVFMGCPG 60
Db 1 MGLPRGLASLLLLQVCWLQCAASEPCRAVFREAEVTLLEAGGAEQEPGALGVFMGCPG 60
QY 61 QEPALFSTNDNDFTVRNGETVQERRSLKERNPLKIPPSKRIILRRHKRWVAPISVPENG 120
Db 61 QEPALFSTNDNDFTVRNGETVQERRSLKERNPLKIPPSKRIILRRHKRWVAPISVPENG 120
QY 121 KGPFPRLNQLKSNKDRDRTKIFYSITGPGADSPPEGVFAVEKETGWLKLNKPLDREEIAK 180
Db 121 KGPFPRLNQLKSNKDRDRTKIFYSITGPGADSPPEGVFAVEKETGWLKLNKPLDREEIAK 180
QY 181 YELFGHAVSENGASVEDPMNISIIVTDNDHDKPKFTQDTRFGSVLEGLVPGTSVMQVTAT 240

Db 181 YELFGHAVSENGASVEDPMNISIIVTDNDHDKPKFTQDTRFGSVLEGLVPGTSVMQVTAT 240
QY 241 DEDDAIYTYNGVAVYSHSQEPKDPHLMFTIHRSTGTISVISGLDREKVPYTLTIOA 300
Db 241 DEDDAIYTYNGVAVYSHSQEPKDPHLMFTIHRSTGTISVISGLDREKVPYTLTIOA 300
QY 301 TMDGDSGTTTAVAVVEILDANDNAPMFDPOKYEAHVPENAVGHVORLTVTDLDPNSP 360
Db 301 TMDGDSGTTTAVAVVEILDANDNAPMFDPOKYEAHVPENAVGHVORLTVTDLDPNSP 360
QY 361 AWRATYILMGDDGDHDTITTHPESNOGILTTTRKGLDPEAKNOHTLYVEVNEAPVVLK 420
Db 361 AWRATYILMGDDGDHDTITTHPESNOGILTTTRKGLDPEAKNOHTLYVEVNEAPVVLK 420
QY 421 PTSTATIVHVEDVNEAPVFPVPSKVVEQEGIPTGEPVCVYTAEDPDKENQISYRILR 480
Db 421 PTSTATIVHVEDVNEAPVFPVPSKVVEQEGIPTGEPVCVYTAEDPDKENQISYRILR 480
QY 481 DPAGWLAMPDSGOVTAVGLTDREDEQFVRNNIYEVNVLAMDNQSPPTTGTTGLLLTLD 540
Db 481 DPAGWLAMPDSGOVTAVGLTDREDEQFVRNNIYEVNVLAMDNQSPPTTGTTGLLLTLD 540
QY 541 VNRHGPVPEPRQITICNOSPVRHVLNITDKLSPHTSPFQAQLTDDSDIYWTAEVNEEGD 600
Db 541 VNRHGPVPEPRQITICNOSPVRHVLNITDKLSPHTSPFQAQLTDDSDIYWTAEVNEEGD 600
QY 601 TVVLSLKKFLKQDQYDVHLSLSDHGKKEQLTVIRATVCDCHGVETCPGPKWGGFLLPVL 660
Db 601 TVVLSLKKFLKQDQYDVHLSLSDHGKKEQLTVIRATVCDCHGVETCPGPKWGGFLLPVL 660
QY 661 GAVLALLFLLVLLVLLVLRKKRKEPLLLPEDDTRDNVYFVYEGEGEEDQDYDITQLHR 720
Db 661 GAVLALLFLLVLLVLLVLRKKRKEPLLLPEDDTRDNVYFVYEGEGEEDQDYDITQLHR 720
QY 721 GLEARPEVLRNDVAPTIITPMYRPRPANDPDEIGNFIENLKAANTDTPAPPYDILLVF 780
Db 721 GLEARPEVLRNDVAPTIITPMYRPRPANDPDEIGNFIENLKAANTDTPAPPYDILLVF 780
QY 781 DYEGSGDAASLSLSSASDQDDYDYLNEWGRPKKLADMYGGEDD 829
Db 781 DYEGSGDAASLSLSSASDQDDYDYLNEWGRPKKLADMYGGEDD 829

RESULT 12
US-10-274-177-16
; Sequence 16, Application US/10274177
; Publication No. US20040038225A1
; GENERAL INFORMATION:
; APPLICANT: MARKOWITZ, Sanford D.
; TITLE OF INVENTION: METHODS FOR CATEGORIZING PATIENTS
; FILE REFERENCE: CWRU-P01-003
; CURRENT APPLICATION NUMBER: US/10/274,177
; CURRENT FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: US/10/229,345
; PRIOR FILING DATE: 2002-08-26
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 16
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-274-177-16

Query Match 87.8%; Score 728; DB 12; Length 829;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 828; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGLPRGLASLLLLQVCWLQCAASEPCRAVFREAEVTLLEAGGAEQEPGALGVFMGCPG 60
Db 1 MGLPRGLASLLLLQVCWLQCAASEPCRAVFREAEVTLLEAGGAEQEPGALGVFMGCPG 60
QY 61 QEPALFSTNDNDFTVRNGETVQERRSLKERNPLKIPPSKRIILRRHKRWVAPISVPENG 120

```
Db      61  QEPALFSTDNDDFTVRNGETVQERSLKERNPLKIFPSKILRRHKDWWVAPISVPENG 120
Qy      121  KGPPFORLNOLKSNKDRDTHKIFYSITGPGADSPPEGVFAVEKETGWLILNKPLDREERIAK 180
Db      121  KGPPFORLNOLKSNKDRDTHKIFYSITGPGADSPPEGVFAVEKETGWLILNKPLDREERIAK 180
Qy      181  YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSVLEGLVPGTSVMQVAT 240
Db      181  YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSVLEGLVPGTSVMQVAT 240
Qy      241  DEDDAIYTYNGVAVSIHSGEPKDPHDLMTTHRSTGTISVSSGLDREKPEYTLTQA 300
Db      241  DEDDAIYTYNGVAVSIHSGEPKDPHDLMTTHRSTGTISVSSGLDREKPEYTLTQA 300
Qy      301  TMDGSGSTTTAVAVVEILDANDNAPMDFPKYEAHVPENAVGEVQRLTVDLDAPNSP 360
Db      301  TMDGSGSTTTAVAVVEILDANDNAPMDFPKYEAHVPENAVGEVQRLTVDLDAPNSP 360
Qy      361  AWRATYLMGDDGDHFTTTHPESNQGLITRKGDLDFEAKNQHTLYVEVTNEAPFVLKL 420
Db      361  AWRATYLMGDDGDHFTTTHPESNQGLITRKGDLDFEAKNQHTLYVEVTNEAPFVLKL 420
Qy      421  PTSTATIVVHVEDVNEAPFVPPSKVVEQGIPTGEPVCVVTAEADPKENOKISYRILR 480
Db      421  PTSTATIVVHVEDVNEAPFVPPSKVVEQGIPTGEPVCVVTAEADPKENOKISYRILR 480
Qy      481  DPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVWVLANMGSPPTTGTGLLLTLD 540
Db      481  DPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVWVLANMGSPPTTGTGLLLTLD 540
Qy      541  VNDHGVPPEPQITICNOSPVRHVLNITDKDLSPTSPFOAQLTDDSDIYWTAEVNEBGD 600
Db      541  VNDHGVPPEPQITICNOSPVRHVLNITDKDLSPTSPFOAQLTDDSDIYWTAEVNEBGD 600
Qy      601  TVVLSLKKFLKQDHYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKGGFIPVL 660
Db      601  TVVLSLKKFLKQDHYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKGGFIPVL 660
Qy      661  GAVLALLFLLVLLVLLVLRKKRKEPILLPEDDTRDNVYVYEGEGGEDDYDITQLHR 720
Db      661  GAVLALLFLLVLLVLLVLRKKRKEPILLPEDDTRDNVYVYEGEGGEDDYDITQLHR 720
Qy      721  GLEARPEVLRNDVAPTIIPTMYRPRPANDPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
Db      721  GLEARPEVLRNDVAPTIIPTMYRPRPANDPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
Qy      781  DYESGSDAASLSLSSASDQDDYDYLNWGSRFKKLADMYGGGEDD 829
Db      781  DYESGSDAASLSLSSASDQDDYDYLNWGSRFKKLADMYGGGEDD 829
```

RESULT 13

```
US-10-301-822-18
; Sequence 18, Application US/10301822
; Public Information No. US20030148410A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: BURGART, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF COLON CANCER
; FILE REFERENCE: MEMO1-029P2RNM
; CURRENT APPLICATION NUMBER: US/10/301.822
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
```

```
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Homo Sapiens
; US-10-301-822-18
```

```
Query Match      87.8%; Score 728; DB 14; Length 829;
Best Local Similarity 99.9%; Pred.No. 0;
Matches 828; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy      1  MGLPRGFLASLLILQVCWLQCAASEPCRAVFREAEVTLGAGAEQEPGQALGVFMGCPG 60
Db      1  MGLPRGFLASLLILQVCWLQCAASEPCRAVFREAEVTLGAGAEQEPGQALGVFMGCPG 60
Qy      61  QEPALFSTDNDDFTVRNGETVQERSLKERNPLKIFPSKILRRHKDWWVAPISVPENG 120
Db      61  QEPALFSTDNDDFTVRNGETVQERSLKERNPLKIFPSKILRRHKDWWVAPISVPENG 120
Qy      121  KGPPFORLNOLKSNKDRDTHKIFYSITGPGADSPPEGVFAVEKETGWLILNKPLDREERIAK 180
Db      121  KGPPFORLNOLKSNKDRDTHKIFYSITGPGADSPPEGVFAVEKETGWLILNKPLDREERIAK 180
Qy      181  YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSVLEGLVPGTSVMQVAT 240
Db      181  YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSVLEGLVPGTSVMQVAT 240
Qy      241  DEDDAIYTYNGVAVSIHSGEPKDPHDLMTTHRSTGTISVSSGLDREKPEYTLTQA 300
Db      241  DEDDAIYTYNGVAVSIHSGEPKDPHDLMTTHRSTGTISVSSGLDREKPEYTLTQA 300
Qy      301  TMDGSGSTTTAVAVVEILDANDNAPMDFPKYEAHVPENAVGEVQRLTVDLDAPNSP 360
Db      301  TMDGSGSTTTAVAVVEILDANDNAPMDFPKYEAHVPENAVGEVQRLTVDLDAPNSP 360
Qy      361  AWRATYLMGDDGDHFTTTHPESNQGLITRKGDLDFEAKNQHTLYVEVTNEAPFVLKL 420
Db      361  AWRATYLMGDDGDHFTTTHPESNQGLITRKGDLDFEAKNQHTLYVEVTNEAPFVLKL 420
Qy      421  PTSTATIVVHVEDVNEAPFVPPSKVVEQGIPTGEPVCVVTAEADPKENOKISYRILR 480
Db      421  PTSTATIVVHVEDVNEAPFVPPSKVVEQGIPTGEPVCVVTAEADPKENOKISYRILR 480
Qy      481  DPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVWVLANMGSPPTTGTGLLLTLD 540
Db      481  DPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVWVLANMGSPPTTGTGLLLTLD 540
Qy      541  VNDHGVPPEPQITICNOSPVRHVLNITDKDLSPTSPFOAQLTDDSDIYWTAEVNEBGD 600
Db      541  VNDHGVPPEPQITICNOSPVRHVLNITDKDLSPTSPFOAQLTDDSDIYWTAEVNEBGD 600
Qy      601  TVVLSLKKFLKQDHYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKGGFIPVL 660
Db      601  TVVLSLKKFLKQDHYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKGGFIPVL 660
Qy      661  GAVLALLFLLVLLVLLVLRKKRKEPILLPEDDTRDNVYVYEGEGGEDDYDITQLHR 720
Db      661  GAVLALLFLLVLLVLLVLRKKRKEPILLPEDDTRDNVYVYEGEGGEDDYDITQLHR 720
Qy      721  GLEARPEVLRNDVAPTIIPTMYRPRPANDPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
Db      721  GLEARPEVLRNDVAPTIIPTMYRPRPANDPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
Qy      781  DYESGSDAASLSLSSASDQDDYDYLNWGSRFKKLADMYGGGEDD 829
Db      781  DYESGSDAASLSLSSASDQDDYDYLNWGSRFKKLADMYGGGEDD 829
```

RESULT 14

US-10-295-027-338

; Sequence 338, Application US/10295027

; Publication No. US20030232350A1

; GENERAL INFORMATION:

; APPLICANT: Afar, Daniel

; APPLICANT: Aziz, Natasha

; APPLICANT: Ginsberg, Wendy M.

; APPLICANT: Gish, Kurt C.

; APPLICANT: Glyne, Richard

; APPLICANT: Hevezi, Peter A.

; APPLICANT: Mack, David H.

; APPLICANT: Murray, Richard

; APPLICANT: Watson, Susan R.

; APPLICANT: Eos Biotechnology, Inc.

; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and

; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer

; FILE REFERENCE: 018501-012500US

; CURRENT APPLICATION NUMBER: US/10/295,027

; CURRENT FILING DATE: 2002-11-13

; PRIOR APPLICATION NUMBER: US 09/663,733

; PRIOR FILING DATE: 2000-09-15

; PRIOR APPLICATION NUMBER: US 60/350,666

; PRIOR FILING DATE: 2001-11-13

; PRIOR APPLICATION NUMBER: US 60/335,394

; PRIOR FILING DATE: 2001-11-15

; PRIOR APPLICATION NUMBER: US 60/332,464

; PRIOR FILING DATE: 2001-11-21

; PRIOR APPLICATION NUMBER: US 60/334,393

; PRIOR FILING DATE: 2001-11-29

; PRIOR APPLICATION NUMBER: US 60/340,376

; PRIOR FILING DATE: 2001-12-14

; PRIOR APPLICATION NUMBER: US 60/347,211

; PRIOR FILING DATE: 2002-01-08

; PRIOR APPLICATION NUMBER: US 60/347,349

; PRIOR FILING DATE: 2002-01-10

; PRIOR APPLICATION NUMBER: US 60/355,250

; PRIOR FILING DATE: 2002-02-08

; PRIOR APPLICATION NUMBER: US 60/356,714

; PRIOR FILING DATE: 2002-02-13

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 1386

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 338

; LENGTH: 829

; TYPE: PRT

; ORGANISM: Homo sapiens

; US-10-295-027-338

Query Match 87.8%; Score 728; DB 15; Length 829;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 828; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	MGLPRGLASLLLLQVWLQCAASEFCRAVFREAEVTLLEAGGAEBPGQALGVFWGCPG	60
DB	1	MGLPRGLASLLLLQVWLQCAASEFCRAVFREAEVTLLEAGGAEBPGQALGVFWGCPG	60
QY	61	QEPALFSTNDNDDFTVNGETVQERRSLKERNPLKIPPSKRIIRRHKRDWVWVAPISVPENG	120
DB	61	QEPALFSTNDNDDFTVNGETVQERRSLKERNPLKIPPSKRIIRRHKRDWVWVAPISVPENG	120
QY	121	KGFPORLQNLKNDKRDYKIFYSITGPGADSPPEGVFAVEKETGWLNLKPLDREELAK	180
DB	121	KGFPORLQNLKNDKRDYKIFYSITGPGADSPPEGVFAVEKETGWLNLKPLDREELAK	180
QY	181	YELFGHAVSNGASVEDPNNISITVDQNDHKPKFTQDTFRGSVLGVLPGTSVMQVAT	240
DB	181	YELFGHAVSNGASVEDPNNISITVDQNDHKPKFTQDTFRGSVLGVLPGTSVMQVAT	240
QY	241	DEDDALTYNGVAVSIHQEPKDPHDLMTFTHRSITGTSVSSGLDREKPEYTLTQA	300
DB	241	DEDDALTYNGVAVSIHQEPKDPHDLMTFTHRSITGTSVSSGLDREKPEYTLTQA	300

QY	301	TDMDGGSTTTAVAVVEILDANDNAPMFPDPOKYEAHVPENAVGHVORLTVTDLDAENSP	360
DB	301	TDMDGGSTTTAVAVVEILDANDNAPMFPDPOKYEAHVPENAVGHVORLTVTDLDAENSP	360
QY	361	AWRATYLLMGDDGDGHFTITTHPSNQIILTRKGLDFEAKNQHTLVVEVNEAPFVLKL	420
DB	361	AWRATYLLMGDDGDGHFTITTHPSNQIILTRKGLDFEAKNQHTLVVEVNEAPFVLKL	420
QY	421	PTSTATIVVHVEDVNEAPFVPPSKVVEQSIGITGSPVCVYTAEDPKENKQKISYILR	480
DB	421	PTSTATIVVHVEDVNEAPFVPPSKVVEQSIGITGSPVCVYTAEDPKENKQKISYILR	480
QY	481	DPAGWLAMPDPSGOVTAAGTLDREDEQFVRNIIYEVWVLMNDGSPPTTGTGTLTLLID	540
DB	481	DPAGWLAMPDPSGOVTAAGTLDREDEQFVRNIIYEVWVLMNDGSPPTTGTGTLTLLID	540
QY	541	VNDHGPVPEPRQITICNQSPVRHLNITDKLSPTSPFQAQLTDDSDIYVTAENNEEGD	600
DB	541	VNDHGPVPEPRQITICNQSPVRHLNITDKLSPTSPFQAQLTDDSDIYVTAENNEEGD	600
QY	601	TVVLSLKKFLKQDQDYDVHLSLSHCKNEQLTVIRATVCDCHGHVETCGPWKGGFILLPVL	660
DB	601	TVVLSLKKFLKQDQDYDVHLSLSHCKNEQLTVIRATVCDCHGHVETCGPWKGGFILLPVL	660
QY	661	GAVLALLFLVLLVLLVLRKKRKEIPELLPEDDDTRDNVFFYVYGGEGGEEQDQYDITQHR	720
DB	661	GAVLALLFLVLLVLLVLRKKRKEIPELLPEDDDTRDNVFFYVYGGEGGEEQDQYDITQHR	720
QY	721	GLEARPEVLENDVAPTIIPTPMYRPRPANDPEIGNFIIENLKAANTDPTAPPYDTLLVF	780
DB	721	GLEARPEVLENDVAPTIIPTPMYRPRPANDPEIGNFIIENLKAANTDPTAPPYDTLLVF	780
QY	781	DYEGSGSDAASLSLTSSASDQDQDYDLNEWSRFPKFLADMTGGGDD	829
DB	781	DYEGSGSDAASLSLTSSASDQDQDYDLNEWSRFPKFLADMTGGGDD	829

RESULT 15

US-10-295-027-1246

; Sequence 1246, Application US/10295027

; Publication No. US20030232350A1

; GENERAL INFORMATION:

; APPLICANT: Afar, Daniel

; APPLICANT: Aziz, Natasha

; APPLICANT: Ginsberg, Wendy M.

; APPLICANT: Gish, Kurt C.

; APPLICANT: Glyne, Richard

; APPLICANT: Hevezi, Peter A.

; APPLICANT: Mack, David H.

; APPLICANT: Murray, Richard

; APPLICANT: Watson, Susan R.

; APPLICANT: Eos Biotechnology, Inc.

; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and

; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer

; FILE REFERENCE: 018501-012500US

; CURRENT APPLICATION NUMBER: US/10/295,027

; CURRENT FILING DATE: 2002-11-13

; PRIOR APPLICATION NUMBER: US 09/663,733

; PRIOR FILING DATE: 2000-09-15

; PRIOR APPLICATION NUMBER: US 60/350,666

; PRIOR FILING DATE: 2001-11-13

; PRIOR APPLICATION NUMBER: US 60/335,394

; PRIOR FILING DATE: 2001-11-15

; PRIOR APPLICATION NUMBER: US 60/332,464

; PRIOR FILING DATE: 2001-11-21

; PRIOR APPLICATION NUMBER: US 60/334,393

; PRIOR FILING DATE: 2001-11-29

; PRIOR APPLICATION NUMBER: US 60/340,376

; PRIOR FILING DATE: 2001-12-14

; PRIOR APPLICATION NUMBER: US 60/347,211

; PRIOR FILING DATE: 2002-01-08

; PRIOR APPLICATION NUMBER: US 60/347,349

; PRIOR FILING DATE: 2002-01-10

Search completed: September 21, 2004, 22:38:15
Job time : 145 secs

; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1246

; LENGTH: 829

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-295-027-1246

Query Match 87.8%; Score 728; DB 15; Length 829;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 828; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	MGLPRGLASLLLLQVCWLCQAASPCRAVPREAEVTLAAGAEOPGQALGKVFMGCPG	60
Db	1	MGLPRGLASLLLLQVCWLCQAASPCRAVPREAEVTLAAGAEOPGQALGKVFMGCPG	60
QY	61	QEPALFSTNDFTVANGETVQERSLKERPNLKIPPSKRIILRRHKRDWVAPISVPENG	120
Db	61	QEPALFSTNDFTVANGETVQERSLKERPNLKIPPSKRIILRRHKRDWVAPISVPENG	120
QY	121	KGFPFQRLNQLKSNKDRDKIYFISITGCGADSPPEGVFAVEKETGNLLINKPLDREETA	180
Db	121	KGFPFQRLNQLKSNKDRDKIYFISITGCGADSPPEGVFAVEKETGNLLINKPLDREETA	180
QY	181	YELFHAVSENGASVEDPMNISIIVTDQNDHKPKFTQDTFRGSVLEGLVPGTSVMQVTAT	240
Db	181	YELFHAVSENGASVEDPMNISIIVTDQNDHKPKFTQDTFRGSVLEGLVPGTSVMQVTAT	240
QY	241	DEDDAIYTYNGVAVSIHQEPKDPHDLMTTHRSTGTISVISGLDRKVPETYLTITQA	300
Db	241	DEDDAIYTYNGVAVSIHQEPKDPHDLMTTHRSTGTISVISGLDRKVPETYLTITQA	300
QY	301	TMDGSGSTTTAVAVEIILDANDNAMPDPQKYEAHVAVENAVGHEVQRLTVTDLAPNSP	360
Db	301	TMDGSGSTTTAVAVEIILDANDNAMPDPQKYEAHVAVENAVGHEVQRLTVTDLAPNSP	360
QY	361	AWRATYLINGGDDGDHFTITTHPESNQILTRKGLDFAKQHTLYEVVTNEAPFVLKL	420
Db	361	AWRATYLINGGDDGDHFTITTHPESNQILTRKGLDFAKQHTLYEVVTNEAPFVLKL	420
QY	421	PSTATIVVHVEDVNEAPVFPVPSKVVEQEGIPTEGVCVYTAEDPDKENOKISYRILR	480
Db	421	PSTATIVVHVEDVNEAPVFPVPSKVVEQEGIPTEGVCVYTAEDPDKENOKISYRILR	480
QY	481	DPAGWLAMPDPSQVTAAGTLDREDEQFVRNNIYEVWVLAMDNAGSPPTTGTGTLTLTID	540
Db	481	DPAGWLAMPDPSQVTAAGTLDREDEQFVRNNIYEVWVLAMDNAGSPPTTGTGTLTLTID	540
QY	541	VNDHGVPEPRQITICNQSPVRHVLNITDKDLSPTSPFOAQLTDDSDIYWTAEVNEEGD	600
Db	541	VNDHGVPEPRQITICNQSPVRHVLNITDKDLSPTSPFOAQLTDDSDIYWTAEVNEEGD	600
QY	601	TVVLSLKKFLKQDTYDVHLSLSDHGNKEOLTIVRAVCDCHGHVETCPGPKGGFLLPVL	660
Db	601	TVVLSLKKFLKQDTYDVHLSLSDHGNKEOLTIVRAVCDCHGHVETCPGPKGGFLLPVL	660
QY	661	GAVLALLFLLVLLVLLVLRKKRIKEPLLPEDDTRDNVIFYEGEGGEDQDYDITQLHR	720
Db	661	GAVLALLFLLVLLVLLVLRKKRIKEPLLPEDDTRDNVIFYEGEGGEDQDYDITQLHR	720
QY	721	GLEARPEVLRNDVAPITIIPTMYRPRPANPDEIGNFIENLKAANTDPTAPPYDTLLVF	780
Db	721	GLEARPEVLRNDVAPITIIPTMYRPRPANPDEIGNFIENLKAANTDPTAPPYDTLLVF	780
QY	781	DYEGSGSDAASLSLSTSSASDQDQDYDYLNEWGSRFKKLADMYGGGEDD	829
Db	781	DYEGSGSDAASLSLSTSSASDQDQDYDYLNEWGSRFKKLADMYGGGEDD	829

This Page Blank (uspto)